ASYMPTOTICS AND COMPUTATIONS FOR APPROXIMATION OF METHOD OF REGULARIZATION ESTIMATORS

A Dissertation

by

SANG-JOON LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2004

Major Subject: Statistics

ASYMPTOTICS AND COMPUTATIONS FOR APPROXIMATION OF METHOD OF REGULARIZATION ESTIMATORS

A Dissertation

by

SANG-JOON LEE

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Randall L. Eubank
(Chair of Committee)

Suojin Wang
(Member)

Joseph D. Ward
(Member)

Thomas E. Wehrly
(Member)

Michael T. Longnecker
(Head of Department)

May 2004

Major Subject: Statistics

ABSTRACT

Asymptotics and Computations for Approximation of Method of Regularization Estimators. (May 2004)

Sang-Joon Lee, B.S., Inha University, Korea;

M.S., Seoul National University, Korea

Chair of Advisory Committee: Dr. Randall L. Eubank

Inverse problems arise in many branches of natural science, medicine and engineering involving the recovery of a whole function given only a finite number of noisy measurements on functionals. Such problems are usually ill-posed, which causes severe difficulties for standard least-squares or maximum likelihood estimation techniques. These problems can be solved by a method of regularization. In this dissertation, we study various problems in the method of regularization. We develop asymptotic properties of the optimal smoothing parameters concerning levels of smoothing for estimating the mean function and an associated inverse function based on Fourier analysis. We present numerical algorithms for an approximated method of regularization estimator computation with linear inequality constraints. New data-driven smoothing parameter selection criteria are proposed in this setting. In addition, we derive a Bayesian credible interval for the approximated method of regularization estimators.

This work is dedicated to my parents, wife and daughters.

ACKNOWLEDGEMENTS

I am indebted to many people for the completion of this dissertation. Most of all, this dissertation would not have been possible without the expert and patient direction of Dr. Randy L. Eubank. His experience and insight have been essential in preparation and completion of this dissertation. His wise guidance was always helpful to me in correcting my foolish errors and writing this dissertation. He taught me not only how to do research but also how to think and solve problems. I shall ever be in debt to him for the influence that he has been in my life. If I had a chance to select my advisor again, I would be his student and I would rely on his wisdom again.

I also would like to thank the other members of my committee, Dr. Suojin Wang, Dr. Thomas E. Wehrly and Dr. Joseph D. Ward for their advice and confidence in achieving my Ph.D. degree. I am also grateful to faculty, staff and colleagues at the Department of Statistics for their help and friendship during my memorable time at Texas A&M University.

This work was supported in part by the former Engineering Imaging Laboratory under the direction of Dr. A. Ted Watson at Chemical Engineering Department. I would like to thank Dr. A. Ted Watson for supporting this research and helping in the development of codes.

I would like to thank my family. Throughout my doctoral studies, I had the loyal and steady support of my loving wife, Sun-Ah. She overlooked my enormous time commitment and the shortage of personal time. She was always confident in my ability to finish this dissertation and kept me strong in the past doctorate years. When I was depressed because of my slow progress in completing this dissertation, she told me that “Feeling exhausted may mean you are already far away from the starting point and close to the goal. Do not forget that you are close to the goal and never lose your aim and intention!" I also thank my lovely daughters Alicia and Sophia for their expressions of love to me.

Because of the unfailing confidence of my brother Sangwoo, sister Hyunjung and sister-in-law Sunyoung and the endless love, unlimited support, encouragement and prayers from my dear parents Jasung and Yongja Lee and parents-in-law Dojoon and Bokhee Kim, I have been able to accomplish my goal. My gratitude toward them is hard to describe and will never fade.

Finally, I would like to express my appreciation and love to God and the Lord Jesus Christ who guides my entire life and my family. I have special thanks for Dr. Chul W. Ahn in the Department of Biostatistics and Internal Medicine at The University of Texas who has been a strong bridge between God and me.

TABLE OF CONTENTS

Page

ABSTRACT . . . iii

DEDICATION . . . iv

ACKNOWLEDGEMENTS . . . v

TABLE OF CONTENTS . . . vii

LIST OF TABLES . . . ix

LIST OF FIGURES . . . x

CHAPTER

I INTRODUCTION . . . 1

1.1 The Problem Posed . . . 1
1.2 Present Status . . . 3
1.3 Procedure . . . 13

II ASYMPTOTICS USING A COSINE FOURIER APPROACH . . . 15

2.1 A Simple Deconvolution Type Problem . . . 15
2.2 Proofs . . . 26

III COMPUTATIONS FOR APPROXIMATION OF METHOD OF REGULARIZATION ESTIMATORS . . . 36

3.1 Introduction . . . 36
3.2 Approximation of MOR Estimators . . . 39
3.3 Development of Hat Matrices . . . 48
3.4 Selection of the Smoothing Parameter . . . 51
3.5 Bayesian Credible Intervals . . . 57
3.6 Proofs . . . 62

IV FINITE SAMPLE STUDIES . . . 68

4.1 Simulation Design . . . 69
4.2 Simulation Results . . . 77
4.3 NMR Experimental Data Analysis . . . 86

CHAPTER Page

V CONCLUSIONS 91
5.1 Summary of Results 91
5.2 Future Research 92

REFERENCES 94

VITA 100

LIST OF TABLES

TABLE Page

1 Empirical properties of the ratio of $\tilde{\lambda}_\mu$ and $\tilde{\lambda}_f$. 78

2 Average (across the design) coverage for the Bayesian credible intervals for the mean function and the inverse function with 1,000 replications. 84

3 The optimal λ values minimizing $\mathrm{GCV}_C(\lambda)$ and $\hat{P}_C(\lambda)$ for sample 1 and sample 2. 86

4 Locations of τ for the modes of the spectra from the estimates of relaxation distribution f. 87

LIST OF FIGURES

FIGURE | Page

1 Selected regions of (α, β) for comparisons of relative convergence rates. 33

2 True inverse functions and $f_k^*(u, \sigma)$. 71

3 Kernel functions. 72

4 Product of true inverse functions and kernel functions. 73

5 True mean functions. 74

6 Fourier coefficients of inverse functions and kernel functions. 75

7 Fourier coefficients of mean functions. 76

8 L_μ, L_f and $GCV_C(\lambda)$ for a simulated data set for f^H and K^H. 80

9 $\mu_{\hat{\lambda}}$ and Bayesian credible interval estimates for a simulated data set for f^H and K^H. 81

10 $f_{\hat{\lambda}}$ and Bayesian credible interval estimates for a simulated data set for f^H and K^H. 82

11 $f_{\hat{\lambda}}$ and Bayesian credible interval estimates for a simulated data set for f^H and K^L. 85

12 Estimates of magnetization function μ and relaxation distribution f for sample 1 and sample 2 from Opalinus clay (Benken, Switzerland). Dots are observed values of magnetization. Solid lines are estimates. 89

13 Estimates of magnetization function μ and relaxation distribution f for sample 3 and sample 4 from water-saturated Texas cream limestone. Dots are measured data. Solid lines are the estimates. . . . 90

CHAPTER I

INTRODUCTION

1.1 The Problem Posed

Inverse problems involve estimation and inference about an unknown function from a finite number of noisy measurements on functionals. Let L be a bounded linear operator that maps f to μ

$$L : f \mapsto \mu, \tag{1.1}$$

where f is an unknown function and μ is a response mean function. Solving the inverse problem is to recover the whole function f from information about μ. Usually these problems are ill-posed. In some cases it is hard to find the unique inverse mapping due to the near singularity of the operator L.

In general, the information about the mean function μ consists of noisy measurements from experimental data. The noise increases the difficulty of recovering f. Often the inverse problem solutions are unstable since the estimate for f could be very sensitive to perturbations in the measurements. In other words, we could encounter a problem of large variation of f from very little perturbations in the observed data. In solving these ill-posed inverse problems, we have severe difficulties using standard least-squares or maximum likelihood estimation techniques.

Solutions to ill-posed inverse problems can be obtained from *the method of regularization* (MOR) (Tikhonov, 1963a, 1963b) with nonparametric regression techniques. To describe the MOR, we consider the standard nonparametric regression

---

The format and style follow that of *Biometrics*.

model first. In fact, MOR can be viewed as a generalization of standard regression smoothing techniques.

A standard nonparametric regression model assumes that we have responses $y_1, \ldots, y_n$ satisfying

$$y_i = \mu(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.2}$$

for $0 \leq t_1 < \cdots < t_n \leq 1$ known design points, $\varepsilon_i,\ i = 1, \ldots, n$, zero mean, uncorrelated random errors with common variance σ^2 and μ an unknown function in

$$\begin{aligned} W_2^m[0,1] = \{\mu : \mu^{(j)} &\text{ is absolutely continuous, } j = 0, \ldots, m-1, \\ &\text{and } \mu^{(m)} \in L_2[0,1]\}. \end{aligned}$$

The mean function μ can be estimated by a smoothing spline μ_λ obtained as the minimizer of

$$n^{-1} \sum_{i=1}^{n} (y_i - g(t_i))^2 + \lambda \int_0^1 (g^{(m)}(t))^2 \, dt, \quad \lambda > 0, \tag{1.3}$$

over all $g \in W_2^m[0,1]$. The smoothing parameter λ in (1.3) governs the trade-off between goodness-of-fit and smoothness.

To generalize (1.2) observe that $\mu(t_i) \equiv L_{t_i}\mu \equiv L_i\mu$ for *evaluation functionals* $L_1, \ldots, L_n$ defined on $W_2^m[0,1]$ when $m \geq 1$. Thus, more generally, let $L_1, \ldots, L_n$ be known, bounded, linear functionals on $W_2^m[0,1]$ with $\mu(t_i) = L_i f,\ i = 1, \ldots, n$, and consider the model

$$y_i = L_i f + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.4}$$

where the ε_i are as in (1.2) and f is some unknown element of $W_2^m[0,1]$. The estimator for f in the generalized model (1.4) can be obtained from MOR. More precisely, the MOR estimator f_λ for f is the minimizer of the criterion

$$n^{-1} \sum_{i=1}^{n} (y_i - L_i g)^2 + \lambda \int_0^1 (g^{(m)}(t))^2 \, dt \tag{1.5}$$

over all $g \in W_2^m[0,1]$. The regularization parameter λ in (1.5) controls the trade-off between goodness-of-fit to the data and smoothness of the estimated inverse function. By selecting a good smoothing parameter λ we can reduce the sensitivity of solutions to perturbations in f.

A typical example of model (1.4) is the Fredholm integral equation of the first kind. It has the form

$$L_t f = \int_0^1 K(t,s) f(s) \, ds, \tag{1.6}$$

where the linear operator L_t is defined by integration against a known bivariate kernel function $K(\cdot,\cdot)$ belonging to $L_2([0,1] \times [0,1])$. In this case, the bounded linear functionals are defined by

$$\begin{aligned} L_i f &\equiv L_{t_i} f \\ &= \mu(t_i) \\ &= \int_0^1 K(t_i,s) f(s) \, ds, \quad i = 1, \ldots, n. \end{aligned}$$

A special case of this formulation corresponds to

$$K(t,s) = K(t-s) \tag{1.7}$$

which gives μ as the convolution of f with the kernel K.

There are numerous topics to be discussed in ill-posed inverse problems including asymptotic studies, computation methods, selection of the smoothing parameter and confidence intervals. Some of the more current developments on these and other related areas will be discussed in the following section.

1.2 Present Status

The MOR estimator of f as the minimizer of (1.5) originated from the work of Tikhonov (1963a, 1963b). Lukas (1980) examined regularization techniques for several types of regularization including Tikhonov regularization, regularization with differential operators, and discrete cases. He provided stabilization conditions for MOR estimators showing that $f_\lambda \to f_0$ as $\lambda \to 0$ and that the sensitivity of f_λ to small perturbations in μ decreases as λ increases for each regularization problem.

Applications of the regularization approach can be found in many branches of modern science. Nychka (1983) applied spline smoothing methodology to an inverse problem involving three-dimensional tumor size distribution in blocks of tissue from measurements on cross-sectional tumor slices. It was observed that parametric regression techniques did not work well in this setting because of the near singularity of the operator in the integral equation. They examined the sampling properties of the MOR estimate by Monte Carlo methods. It was also pointed out that the eigenvalues of the integral kernel provide information on ill posedness. Nychka (1984) examined the same problem focusing on the selection of the smoothing parameter. Other examples are provided by the study of porous media by the nuclear magnetic resonance technique (Liaw, Kullkarni, Chen, and Watson, 1996), crystallography (Grunbaum, 1975), and geophysics (Aki and Richards, 1980; Bolt, 1980).

We briefly review some necessary functional analysis concepts and notation to summarize asymptotic developments associated with MOR estimators. A *normed* space is a vector space with a norm defined on it and a *Banach* space is a complete normed space. An inner product space is a vector space with an inner product $< \cdot, \cdot >$ defined on it. It is very easy to verify that inner product spaces are normed spaces. A *Hilbert* space is a complete inner product space (complete in the metric defined by the inner product). Most of the asymptotic properties of MOR estimators have been examined in Hilbert space settings.

Let $\mathcal{F}$ be a real Hilbert space with norm $\|\cdot\|$ and inner product $< \cdot, \cdot >$. Consider a functional of a quadratic form $\frac{1}{2} < f, Wf >$, where W is a nonnegative operator on $\mathcal{F}$. Suppose $\mathbf{y}_n = (y_1, \ldots, y_n)^T$ and L_n is an operator from $\mathcal{F}$ to R^n so that $L_n f$ is an n-vector. In this setting the MOR estimator f_λ is obtained by minimizing

$$(\mathbf{y}_n - L_n f)^T(\mathbf{y}_n - L_n f) + \lambda < f, Wf > . \tag{1.8}$$

Let $U_n = L_n^* L_n$ and set $G_{n\lambda} = (\lambda W + U_n)^{-1}$, where L_n^* is the adjoint of L_n. Assuming $G_{n\lambda}^{-1}$ exists,

$$f_\lambda = G_{n\lambda}^{-1} L_n^* \mathbf{y}_n, \qquad 0 < \lambda < \infty.$$

To study the asymptotic behavior of f_λ assume there is a positive sequence $\{a_n\}$ and a fixed bounded linear operator U such that $a_n L_n^* L_n \to U$ when $n \to \infty$. Cox (1988) investigated asymptotic properties of f_λ in a convenient family of norms which are defined in terms of U and W. Assume U is a compact positive definite operator on $\mathcal{F}$ and W is a nonnegative definite linear operator on $\mathcal{F}$. Then there are sequences $\{\phi_\nu : \nu = 1, 2, \ldots\}$ and $\{\gamma_\nu : \nu = 1, 2, \ldots\}$ of eigenfunctions and eigenvalues, respectively, satisfying

$$< \phi_\eta, U\phi_\nu >= \delta_{\nu\eta} \text{ and } < \phi_\eta, W\phi_\nu >= \gamma_\nu \delta_{\nu\eta}, \tag{1.9}$$

for all pairs ν, η of positive integers, where $\delta_{\nu\eta}$ is a Kronecker's delta.

For $\rho \in R$, let

$$\|f\|_\rho = \Big[\sum_{\nu=1}^{\infty} (1 + \gamma_\nu^\rho) < f, U\phi_\nu >^2 \Big]^{1/2} \tag{1.10}$$

and let $\mathcal{F}_\rho$ be the normed linear space obtained by completing $\{f \in \mathcal{F} : \|f\|_\rho < \infty\}$. Then, $\mathcal{F}_\rho$ is a Hilbert space with inner product

$$< f, \zeta >_\rho = \sum_{\nu=1}^{\infty} (1 + \gamma_\nu^\rho) < f, U\phi_\nu >< \zeta, U\phi_\nu > . \tag{1.11}$$

Cox (1988) developed rates of convergence for MOR estimators in the family of norms provided by (1.10). The consistency and the rate of convergence of the MOR estimator are examined through the mean squared error $\mathrm{E}||f - f_\lambda||_\rho^2$ which is decomposed into bias and variance components. For the MOR estimator f_λ, it is shown that the bias $||f - \mathrm{E}f_\lambda||_\rho^2$ tends to 0 or has an upper bound depending under certain conditions. Under some regularity conditions concerning decay rates for the eigenvalues of L_n, the limits of the variance term $\mathrm{E}||f_\lambda - \mathrm{E}f_\lambda||_\rho^2$ are developed when $\lambda \to 0$.

Nychka and Cox (1989) extended the work by Cox (1988) to the general case where the decay rates of the eigenvalues of L_n are not known. Under the Hilbert space norm defined in (1.10) and the associated inner product in (1.11), let m be the dimension of the null space of the operator W in (1.9) and write the n-vector of errors as $\varepsilon_n = \mathbf{y}_n - L_n f$. The assumptions imposed are

(1) For $\eta \in \mathbf{R}^n$, $\mathrm{E}\varepsilon_n = \mathbf{0}$ and there are constants $\{S_n\} \subseteq (0, \infty)$ such that $\mathrm{E} < \eta, \varepsilon_n >^2 = O(S_n \|\eta\|^2)$ and $S_n \|\eta\|^2 = O(\mathrm{E} < \eta, \varepsilon_n >^2)$,

(2) There are values $0 < r \leq q < \infty$ such that $j^r = O(\gamma_j)$ and $\gamma_j = O(j^q)$, where γ_j is as in (1.9) and (1.10),

(3) There are real numbers $s \in (0, 1 - 1/r)$, $\{\rho_1, \rho_2, \ldots, \rho_j\} \subset [0, s]$ and $d_n \subseteq [0, \infty)$ with $d_n \to 0$ such that for all $f_1, f_2 \in \mathcal{F}$,

$$| < f_1, U f_2 > - < L_n f_1, L_n f_2 > | \leq d_n \sum_{i=1}^{j} ||f_1||_{\rho_i} ||f_2||_{s - \rho_i}.$$

Nychka and Cox (1989) showed that under assumptions (1)-(3) and regularity conditions on s, r, ρ, β and d_n, for $0 \leq \rho < 2 - s - 1/r$, $\rho < \beta \leq \rho + 2$, and $\lambda_n \to 0$ as $n \to \infty$,

$$\mathrm{E}||f_\lambda - f||_\rho^2 = O\left(\min\{1, \lambda^{\beta - \rho}\} ||f||_\beta^2 + S_n \{ \sum_{j > m} \gamma_j^\rho (1 + \lambda \gamma_j)^{-2} + m \} \right)$$

uniformly for $\lambda \in [\lambda_n, \infty)$.

Further extensions of the asymptotic developments for MOR estimator were made by Cox and O'Sullivan (1990). Cox and O'Sullivan (1990) developed the first order asymptotic properties for MOR estimators in the cases of log-density estimation, log-hazard estimation, nonparametric logistic regression, and more general nonlinear inverse problem examples. Their approach is based on a functional Taylor series expansion technique. One of the main asymptotic results is that the optimal rate of convergence of an MOR estimator is

$$||f_{\lambda^*} - f||_\rho^2 = O_p\big(n^{-2m(p-\rho)/(2m+1)}\big),$$

where ρ is as in (1.10) and the "optimal" smoothing parameter λ^* is of order $n^{-2m/(2m+1)}$. Similar asymptotic work based on kernel density estimators for deconvolution problems can be found in Stefanski (1990) and Fan (1991).

Wahba (1980) developed practical computational methods for MOR estimator based on the approximation

$$f(s) \doteq \sum_{j=1}^{N} c_j B_j(s), \tag{1.12}$$

where the $B_j(s), j = 1, \ldots, N$, are B-spline basis functions. For B-spline functions of degree $2m-1$, we have

$$\begin{aligned} \Sigma &= \Big\{ \int_0^1 B_j^{(m)}(s) B_k^{(m)}(s) ds \Big\}_{j,k=1,\ldots,N} && (1.13) \\ &= \Gamma S \Gamma^T, && (1.14) \end{aligned}$$

where Σ is a matrix of rank $N-m$, Γ is the $N \times (N-m)$ matrix whose columns are eigenvectors of Σ associated with its non zero eigenvalues, S is the $(N-m) \times (N-m)$ diagonal matrix of non zero eigenvalues and Δ is the $N \times m$ matrix whose columns are the eigenvectors associated with the zero eigenvalues. The decomposition of Σ in (1.14) makes it possible for the coefficient vector $\mathbf{c} = (c_1, c_2, \ldots, c_N)^T$ to be decomposed into elements of the null space of Σ and elements in the null space of $\Sigma^\perp$ such that

$$\mathbf{c} = \Gamma S^{-1/2} \Lambda + \Delta \mathbf{d},$$

for some $\Lambda = (\gamma_1, \ldots, \gamma_{N-m})^T$, and $\mathbf{d} = (d_1, \ldots, d_m)^T$. Define the $n \times N$ matrix $\mathbf{L} \equiv \{\mathbf{l}_1, \ldots, \mathbf{l}_n\}^T$, where $\mathbf{l}_i = \big(\int_0^1 K(t_i, s)B_1(s)ds, \ldots, \int_0^1 K(t_i, s)B_N(s)ds \big)^T$ and let $\mathbf{y} = (y_1, \ldots, y_n)^T$. Then, we observe that the Tikhonov regularization criterion in (1.3) can be approximated by

$$n^{-1}||\mathbf{y} - \mathbf{L}\mathbf{c}||^2 + \lambda \mathbf{c}^T \Sigma \mathbf{c} = n^{-1}||\mathbf{y} - \mathbf{L}\Gamma S^{-1/2}\Lambda - \mathbf{L}\Delta\mathbf{d}||^2 + \lambda \Lambda^T \Lambda. \tag{1.15}$$

Let $P = \mathbf{L}\Delta\big[(\mathbf{L}\Delta)^T(\mathbf{L}\Delta)\big]^{-1}(\mathbf{L}\Delta)^T$ and $W = (I - P)\mathbf{L}\Gamma S^{-1/2}$, where I is an identity matrix and let V and D be orthogonal and diagonal matrices, respectively, such that $W^T W = V D V^T$. Then, by differentiating the criterion (1.15) with respect to $\mathbf{d}$ and Λ, one finds that the minimizer of (1.15) is

$$\mathbf{c}_\lambda = \Gamma S^{-1/2}\Lambda^* + \Delta \mathbf{d}^*,$$

where

$$\mathbf{d}^* = \big[(\mathbf{L}\Delta)^T(\mathbf{L}\Delta)\big]^{-1}(\mathbf{L}\Delta)^T(\mathbf{y} - \mathbf{L}\Gamma S^{-1/2}\Lambda)$$

and

$$\Lambda^* = V(D + n\lambda I)^{-1}V^T W^T \mathbf{y}.$$

Data-adaptive selection of the smoothing parameter λ is a very important problem to be considered in practice. The selection of a smoothing parameter plays an important role in estimation. In estimation of the mean function, for example, if $\lambda = 0$ then the solution is a spline function interpolating the observed data. On the other hand, when $\lambda = \infty$ the minimization problem is to minimize residual sum-of-squares with $\int_0^1 (g^{(m)}(t))^2 \, dt = 0$ in (1.3), which results in polynomial regression. Excluding these extreme values, very large values of λ give an estimated function that is generally too smooth and looses a large proportion of information in the data. On the other hand, when λ is too small the estimated function tends to be rough or wiggly.

In the MOR context, the goodness-of-fit term corresponds to $n^{-1}\sum_{i=1}^{n}(y_i - L_i g)^2$ in (1.5) since there is no directly related data for the inverse function f. Using this assessment of fit classical data-driven smoothing parameter selection criteria, specifically, *cross-validation* (CV) (Allen, 1974), *generalized cross-validation* (GCV) (Craven and Wahba, 1979), *unbiased risk estimation* (Mallows, 1973) and *one-sided cross-validation* (Hart and Yi, 1998) can be employed for the MOR estimator.

Rice (1986) demonstrated that a reasonable choice of the smoothing parameter for making mean squared error small for estimating μ may not be reasonable in terms of the estimation error incurred for f, and vice versa. O'Sullivan (1986) overviewed various issues in MOR estimation and the solution to ill-posed inverse problems with an extension of CV and related smoothing parameter selection criteria that might be useful for the MOR case. However Wahba (1986) pointed out that the smoothing parameter selection criteria proposed by O'Sullivan (1986) can be ill-posed.

One of the most important issues regarding selection of smoothing parameters in the MOR context is when the classical CV or related criteria can provide a good estimate of the smoothing parameter for estimating the function f. In this regard, a study of the optimal smoothing parameters for estimating both f and μ has been considered by Wahba (1990) for the special case of convolution. Specifically, she assumed that

$$\mu(t) = \int_0^1 K(t-s)f(s)\,ds \tag{1.16}$$

with periodic kernels and periodic smoothing splines being used to investigate the large sample properties of μ_λ and f_λ. The Fourier series expansions of μ, K and f in (1.16) are

$$\mu(t) = \sum_{j=1}^{\infty} 2\mu_j \cos(2\pi jt), \quad \text{with} \quad \mu_j = \int_0^1 \cos(2\pi jt)\mu(t)\,dt,$$

$$K(t) = \sum_{j=1}^{\infty} 2k_j \cos(2\pi jt), \quad \text{with} \quad k_j = \int_0^1 \cos(2\pi jt)K(t)\ dt,$$
$$f(t) = \sum_{j=1}^{\infty} 2f_j \cos(2\pi jt), \quad \text{with} \quad f_j = \int_0^1 \cos(2\pi jt)f(t)\ dt.$$

These are employed for asymptotic studies to show convergence properties of the optimal λ, minimizing either the integrated domain risk,

$$\mathrm{IDR}(\lambda) = \int_0^1 \mathrm{E}(f(t) - f_\lambda(t))^2\ dt \tag{1.17}$$

or the integrated range risk,

$$\mathrm{IRR}(\lambda) = \int_0^1 \mathrm{E}(\mu(t) - \mu_\lambda(t))^2\ dt \tag{1.18}$$

as global measures of the performance of the estimators of f and μ. The Fourier coefficients are assumed to decay algebraically with

$$f_j \sim j^{-\alpha}, \quad k_j \sim j^{-\beta} \tag{1.19}$$

for

$$\alpha, \beta \geq 0, \quad j = 1, 2, \ldots$$

Let λ_μ and λ_f be the minimizers of (1.18) and (1.17), respectively, and take $\|f\|^2 = \int_0^1 (f^{(m)}(t))^2\ dt$. Then, it is shown that as $n \to \infty$,

1. when $2m + \frac{1}{2} \geq \alpha \geq \frac{1}{2}$ and $m > \frac{1}{4}$,

$$\lambda_\mu \asymp \lambda_f = O(n^{-\frac{m+\beta}{\alpha+\beta}}), \tag{1.20}$$

2. when $\alpha > 2m + \frac{1}{2}$ and $m > \frac{1}{4}$,

$$\lambda_\mu \asymp \lambda_f = O(n^{-\frac{m+\beta}{\alpha+\beta}}), \tag{1.21}$$

3. and when $\beta > \alpha - (2m + \frac{1}{2})$ or $\beta \le \alpha - (2m + \frac{1}{2})$,

$$\lambda_\mu = o(\lambda_f). \tag{1.22}$$

However Parseval's relation reveals that

$$\begin{aligned}\int_0^1 (f^{(m)}(t))^2 \, dt &= 2^{-1} \sum_{j=1}^{\infty} (2\pi)^{2m} j^{2m} f_j^2 \\ &\sim 2^{-1} \sum_{j=1}^{\infty} (2\pi)^{2m} j^{2m-2\alpha}.\end{aligned}$$

Hence, $f^{(m)}$ does not belong to $L_2[0,1]$ when $2m - 2\alpha \ge -1$.

When we have experimental data, sometimes it is necessary to place some constraints on the function f to produce physically valid estimators for the problem at hand. Wahba (1980) suggested an algorithm for MOR estimators, CV and, GCV for inequality constrained MOR estimation problems based on the active constraint set for inequality constraints of the form

$$N_j f \le r_j, \quad j = 1, \ldots, k,$$

with the N_j's are bounded linear functionals. It finds a solution iteratively adding in one of the most violated constraints until the solution satisfies all the constraints. Let $f_{\lambda c}$ be the minimizer of (1.5) satisfying all the constraints, and let $f_{\lambda c}^{[q]}$ be the minimizer of

$$n^{-1} \sum_{i=1, i \ne q}^{n} (y_i - g(t_i))^2 + \lambda \int_0^1 (g^{(m)}(t))^2 \, dt, \quad \lambda > 0 \tag{1.23}$$

satisfying the constraints. Let $a_{qq}^*(\lambda, \delta)$ be the differential influence of the qth data point on the qth predicted value, when the qth data point is perturbed by the amount of δ, specifically,

$$a_{qq}^*(\lambda, \delta) = \{L_q f_{\lambda c}[y_1, \ldots, y_{q-1}, y_q + \delta, y_{q+1}, \ldots, y_n] - L_q f_{\lambda c}[y_1, \ldots, y_n]\}/\delta,$$

where the notation $f_{\lambda c}[y_1, \ldots, y_n]$ represents that $f_{\lambda c}$ is the constrained minimizer of (1.23) given the data $y_1, \ldots, y_n$. In this setting, it is shown that

$$\begin{aligned} \mathrm{CV}(\lambda) &= \frac{1}{n}\sum_{i=1}^{n}\left[y_i - L_i f_{\lambda c}^{[i]}\right]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\left[y_i - L_i f_{\lambda c}\right]^2}{1 - a_{ii}^*(\lambda, (L_i f_{\lambda c}^{[i]} - y_i))}, \end{aligned}$$

and

$$\mathrm{GCV}(\lambda) = \frac{(1/n)\sum_{i=1}^{n}\left[y_i - L_i f_{\lambda c}\right]^2}{\left[1 - (1/n)\sum_{i=1}^{n} a_{ii}^*(\lambda, (L_i f_{\lambda c}^{[i]} - y_i))\right]^2}.$$

Wahba (1982) extended CV and GCV to the MOR problem with constraints in a closed convex set with a detailed computational algorithm for the CV and GCV criteria along with applications in meteorology and medicine. Villalobos and Wahba (1987) adapted the method of GCV for constrained problems for estimating a good value of the smoothing parameter. Their computation of the GCV estimate of the smoothing parameter and the constrained spline is based on the active set algorithm of method 3 of Gill, Gould, Murray, Saunders, and Wright (1982).

It is known (e.g., Eubank (1999), Section 5.6 and references therein) that under a particular Bayesian model the smoothing spline fitted values $\mu_\lambda(t_1), \ldots, \mu_\lambda(t_n)$ are the posterior means of $\mu(t_1), \ldots, \mu(t_n)$. This makes it possible to construct a Bayesian credible interval of the form

$$P(|\mu(t_i) - \mu_\lambda(t_i)| \leq z_{\alpha/2}\sigma\sqrt{s_i}) = 1 - \alpha, \tag{1.24}$$

where s_i is a diagonal element of the smoothing spline hat matrix that maps responses $\{y_1, \ldots, y_n\}$ to fitted values $\{\mu_\lambda(t_1), \ldots, \mu_\lambda(t_n)\}$. The frequentist properties of these intervals have been studied by Wahba (1990) and Nychka (1988, 1990). In particular, Nychka (1988) showed that when the coverage probabilities for point-wise intervals are averaged across the observational data, the average coverage probability is close to the point-wise nominal level with CV being used for smoothing parameter selection. Using GCV, Nychka (1990) also showed that the estimates of the average posterior variance converges in probability to a quantity proportional to the expected average squared error.

1.3 Procedure

Our large sample development will focus on a particular MOR estimator for the case where the linear functionals are defined by a convolution type operator. In Chapter II, we present estimators for the mean function and inverse function and discover the asymptotic relation between the optimal smoothing parameters concerning the levels of smoothing for estimating the mean function μ and the associated inverse function f.

Chapter III focuses on practical implementation of the MOR method. We describe an estimation problem concerning the relaxation distribution from nuclear magnetic resonance (NMR) experiments. We present methods for the computation of MOR estimators in the presence of constraints and criteria for choosing the smoothing parameter. The motivations for CV and related methods for estimating the mean function and issues and methodology changes in the MOR context are also presented. Then we discuss Bayesian credible intervals for the mean function and the associated inverse function.

In Chapter IV, we perform finite sample studies for our constrained MOR estimator with nonnegativity constraints. We investigate the nature of these ill-posed inverse problems based on Fourier series implementation. The performance of Bayesian credible intervals based on empirical point-wise coverage probability are investigated. We apply our methods to both simulated and experimental data sets to evaluate their performance. Finally, in Chapter V, we summarize the results presented in this dissertation and discuss their implications with suggestions for future research.

CHAPTER II

ASYMPTOTICS USING A COSINE FOURIER APPROACH

2.1 A Simple Deconvolution Type Problem

In this section we investigate the properties of a particular MOR estimator for the case where the linear functionals are defined by a convolution type operator of the form

$$L_t f = \int_0^1 \frac{1}{2}\left[K(t-s)+K(t+s)\right] f(s)\ ds, \tag{2.1}$$

where $t \in [0,1]$, the kernel K is assumed to be symmetric on $[-1,1]$ extended periodically to $[-1,2]$ with $K \in C^1[-1,2]$. The model is

$$\begin{aligned} y_i &= L_i f + \varepsilon_i \\ &= L_{t_i} f + \varepsilon_i \\ &= \mu(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \end{aligned} \tag{2.2}$$

with $t_i = (2i-1)/2n,\ i = 1, \ldots, n$, and the ε_i are as in (1.2).

Fourier series provides a means of approximating a function $f \in L_2$. In fact the Fourier series expansion of f is identical to f in the sense of L_2. For $m \geq 1$, the Fourier series expansion of f converges to f point-wise. We employ the following cosine Fourier series expansions and coefficients of f and μ:

$$f(t) = f_0 + \sum_{j=1}^{\infty} f_j \sqrt{2}\cos(j\pi t),$$

where

$$f_0 = \int_0^1 f(t)\ dt,$$

$$f_j = \int_0^1 f(t)\sqrt{2}\cos(j\pi t)\ dt,\ j = 1, 2, \ldots,$$

and

$$\mu(t) = \mu_0 + \sum_{j=1}^{\infty} \mu_j \sqrt{2} \cos(j\pi t),$$

where

$$\mu_0 = \int_0^1 \mu(t)\ dt \tag{2.3}$$
$$= \int_0^1 L_t f\ dt, \tag{2.4}$$

$$\mu_j = \int_0^1 \mu(t)\sqrt{2}\cos(j\pi t)\ dt \tag{2.5}$$
$$= \int_0^1 L_t f \sqrt{2}\cos(j\pi t)\ dt,\ \ j = 1, 2, \ldots \tag{2.6}$$

The following proposition states that the cosine Fourier series expansion of the kernel function in $[0, 1]$ holds uniformly over $[-1, 2]$.

Proposition 2.1. *Let K be symmetric on $[-1, 1]$ extended periodically to $[-1, 2]$ with $K \in C^1[-1, 2]$. Then,*

$$K(u) = k_0 + \sum_{j=1}^{\infty} k_j \sqrt{2}\cos(j\pi u) \tag{2.7}$$

uniformly in $[-1, 2]$ with

$$k_0 = \int_0^1 K(u)\ du, \tag{2.8}$$

and

$$k_j = \int_0^1 K(u)\sqrt{2}\cos(j\pi u)\ du, \quad j = 1, 2, \ldots \tag{2.9}$$

Using Proposition 2.1, we develop a cosine version of the convolution theorem that holds for the simple deconvolution type problem in (2.1) and allows us to connect the Fourier coefficients for f to those for μ.

Theorem 2.1. *For the convolution type operator L_t as in (2.1), and K as in Proposition 2.1,*

$$\mu_0 = k_0 f_0, \tag{2.10}$$

$$\mu_j = \frac{1}{\sqrt{2}} k_j f_j, \quad j = 1, 2, \ldots \tag{2.11}$$

To estimate f we will use an approximation to the minimizer of

$$n^{-1} \sum_{i=1}^{n} (y_i - L_i g)^2 + \lambda \int_0^1 (g'(t))^2 \, dt \tag{2.12}$$

over all $g \in W_2^1[0,1]$ that we obtain from a Fourier analysis argument. From Fourier analysis we know that any function in $W_2^m[0,1]$ and, in particular, the minimizer of our MOR criterion can be expressed as

$$g(\cdot) = g_0 + \sum_{j=1}^{\infty} g_j \sqrt{2} \cos(j\pi(\cdot)).$$

We will restrict attention to a finite dimensional approximation of the form

$$\tilde{g}(\cdot) = g_0 + \sum_{j=1}^{n} g_j \sqrt{2} \cos(j\pi(\cdot)).$$

Now

$$L_{t_i} \tilde{g}(\cdot) = k_0 g_0 + \sum_{j=1}^{n} \frac{k_j g_j}{\sqrt{2}} \sqrt{2} \cos(j\pi t_i)$$

and the cosine functions are orthogonal across our uniform design. So, our criterion reduces to a minimization of

$$n^{-1}(\mathbf{y} - \mathbf{XKg})^T(\mathbf{y} - \mathbf{XKg}) + \lambda \mathbf{g}^T \mathbf{D} \mathbf{g}, \tag{2.13}$$

where

$$\mathbf{X} = \left[1 | \{\sqrt{2}\cos(j\pi t_i)\}_{i=1,n,j=1,n-1}\right],$$

$$\mathbf{K} = (\sqrt{2})^{-1} \mathrm{diag}(\sqrt{2} k_0, k_1, \ldots, k_{n-1}),$$

$$\mathbf{D} = \mathrm{diag}(0, \pi^2, \ldots, ((n-1)\pi)^2),$$

$$\mathbf{y} = (y_1, \ldots, y_n)^T,$$

and

$$\mathbf{g} = (g_0, \ldots, g_{n-1})^T.$$

Elementary calculus gives us estimators of μ and f; namely, with $\tilde{y}_0 = \bar{y}$, the response average, and $\tilde{y}_j = n^{-1} \sum_{i=1}^{n} y_i \sqrt{2} \cos(j\pi t_i),\ j = 1, \ldots, n-1$, the sample cosine Fourier coefficients, we have

$$\begin{aligned} \mu_\lambda(t) &= \tilde{y}_0 + \sum_{j=1}^{n-1} \frac{k_j^2 \tilde{y}_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2} \cos(j\pi t) \\ &= \mu_{\lambda 0} + \sum_{j=1}^{n-1} \mu_{\lambda j} \sqrt{2} \cos(j\pi t), \end{aligned} \tag{2.14}$$

where $\mu_{\lambda 0} = \tilde{y}_0$ and $\mu_{\lambda j} = \frac{k_j^2 \tilde{y}_j}{k_j^2 + 2\lambda(j\pi)^2},\ j = 1, \ldots, n-1$, and (for $k_0 \neq 0$)

$$\begin{aligned} f_\lambda(t) &= \frac{\tilde{y}_0}{k_0} + \sum_{j=1}^{n-1} \frac{\sqrt{2} k_j \tilde{y}_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2} \cos(j\pi t) \\ &= f_{\lambda 0} + \sum_{j=1}^{n-1} f_{\lambda j} \sqrt{2} \cos(j\pi t), \end{aligned} \tag{2.15}$$

where $f_{\lambda 0} = \frac{\tilde{y}_0}{k_0}$ and $f_{\lambda j} = \frac{\sqrt{2} k_j \tilde{y}_j}{k_j^2 + 2\lambda(j\pi)^2},\ j = 1, \ldots, n-1$. This follows from the fact that

$$\mathbf{f}_\lambda = \left[\mathbf{K}^2 + \lambda \mathbf{D}\right]^{-1} \mathbf{K}^T \left(n^{-1} \mathbf{X}^T \mathbf{y}\right)$$

is the minimizer of (2.13) as a function of $\mathbf{g}$.

The following corollary indicates how the estimators in (2.14) and (2.15) weight the responses.

Corollary 2.1. *The estimators in (2.14) and (2.15) can be expressed as the following linear combinations of responses:*

$$\mu_\lambda(t) = \frac{1}{n} \sum_{i=1}^{n} y_i w_\mu(t, t_i),$$

and

$$f_\lambda(t) = \frac{1}{n} \sum_{i=1}^{n} y_i w_f(t, t_i),$$

where

$$w_\mu(t,t_i) = 1 + \sum_{j=1}^{n-1} \frac{k_j^2}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2}\cos(j\pi t)\sqrt{2}\cos(j\pi t_i),$$

and

$$w_f(t,t_i) = \frac{1}{k_0} + \sum_{j=1}^{n-1} \frac{\sqrt{2}k_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2}\cos(j\pi t)\sqrt{2}\cos(j\pi t_i),$$

for $k_0 > 0$.

To assess the performance of our estimators of μ and f we consider the integrated domain risk,

$$\mathrm{IDR}(\lambda) = \int_0^1 \mathrm{E}(f(t) - f_\lambda(t))^2 \; dt \tag{2.16}$$

and the integrated range risk,

$$\mathrm{IRR}(\lambda) = \int_0^1 \mathrm{E}(\mu(t) - \mu_\lambda(t))^2 \; dt \tag{2.17}$$

as global measures of the mean squared errors of the estimators of f and μ, respectively. Here we investigate the optimal λ's asymptotic behavior for our particular MOR estimator and develop some asymptotic relations between smoothing parameters minimizing $\mathrm{IDR}(\lambda)$ and $\mathrm{IRR}(\lambda)$.

First observe that

$$\begin{aligned}
\mathrm{IDR}(\lambda) &= \int_0^1 \mathrm{E}(f(t) - f_\lambda(t))^2 \; dt \\
&= \int_0^1 \mathrm{E}(f(t) - \mathrm{E}f_\lambda(t) + \mathrm{E}f_\lambda(t) - f_\lambda(t))^2 \; dt \\
&= \int_0^1 (f(t) - \mathrm{E}f_\lambda(t))^2 \; dt + \int_0^1 \mathrm{Var} f_\lambda(t) \; dt \\
&= \sum_{j=0}^{\infty} (f_j - \mathrm{E}f_{\lambda j})^2 + \sum_{j=0}^{n-1} \mathrm{Var} f_{\lambda j} \\
&= \sum_{j=0}^{n-1} (f_j - \mathrm{E}f_{\lambda j})^2 + \sum_{j=n}^{\infty} f_j^2 + \sum_{j=0}^{n-1} \mathrm{Var} f_{\lambda j}.
\end{aligned}$$

Similarly we have

$$\begin{aligned}\mathrm{IRR}(\lambda) &= \int_0^1 \mathrm{E}(\mu(t) - \mu_\lambda(t))^2 \ dt \\ &= \sum_{j=0}^{n-1} (\mu_j - \ \mathrm{E}\mu_{\lambda j})^2 + \sum_{j=n}^{\infty} \mu_j^2 + \sum_{j=0}^{n-1} \ \mathrm{Var}\mu_{\lambda j}.\end{aligned}$$

Now note that

$$\begin{aligned}\mathrm{E}(\tilde{y}_0) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) && (2.18)\\ &= \frac{1}{n}\sum_{i=1}^{n} L_i f && (2.19)\\ &\doteq \mu_0 && (2.20)\\ &= k_0 f_0 && (2.21)\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}(\tilde{y}_j) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} y_i \sqrt{2}\cos(j\pi t_i)\right) && (2.22)\\ &= \frac{1}{n}\sum_{i=1}^{n} L_i f \sqrt{2}\cos(j\pi t_i) && (2.23)\\ &\doteq \mu_j && (2.24)\\ &= \frac{1}{\sqrt{2}} k_j f_j, \quad j = 1, \ldots, n-1. && (2.25)\end{aligned}$$

Let

$$W_n(t) = \frac{\text{number of } t_i \leq t}{n}.$$

Using the same ideas as in Eubank (1999) and integration by parts, the quadrature errors due to replacing the $L_2[0,1]$ Fourier coefficients with their discrete approximations in (2.19) and (2.23) are

$$r_0 = \left| \ \frac{1}{n}\sum_{i=1}^{n} L_i f - \int_0^1 L_t f \ dt \ \right| \qquad (2.26)$$

$$= \left| \int_0^1 L_t f \; d(W_n(t) - t) \right| \tag{2.27}$$

$$= O(n^{-1}) \|L_t f\| \tag{2.28}$$

and

$$r_j = \left| \frac{1}{n} \sum_{i=1}^{n} L_i f \sqrt{2} \cos(j\pi t_i) - \int_0^1 L_t f \sqrt{2} \cos(j\pi t) \; dt \right| \tag{2.29}$$

$$= \left| \int_0^1 L_t f \sqrt{2} \cos(j\pi t) \; d(W_n(t) - t) \right| \tag{2.30}$$

$$\leq \left| \sqrt{2} \cos(j\pi t) \right| \left| \int_0^1 L_t f \; d(W_n(t) - t) \right| \tag{2.31}$$

$$= O(n^{-1}) \|L_t f\|, \qquad j = 1, \ldots, n-1.$$

Thus, for a function $\mu(\cdot)$ in W_2^1, r_0 in (2.26) and r_j, $j = 1, \ldots, n-1$ in (2.29) will generally be negligible and we will ignore quadrature error in what follows.

Finally we see that the $\tilde{y}_j$ are uncorrelated in that

$$\mathrm{Var}(\tilde{y}_0) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} y_i \right) = \frac{\sigma^2}{n},$$

$$\begin{aligned}
\mathrm{Cov}(\tilde{y}_0, \tilde{y}_j) &= \mathrm{Cov}\left(\frac{1}{n} \sum_{i=1}^{n} y_i, \frac{1}{n} \sum_{i=1}^{n} y_i \sqrt{2} \cos(j\pi t_i) \right) \\
&= \mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} y_i - \mathrm{E}(\tilde{y}_0) \right) \left(\frac{1}{n} \sum_{i=1}^{n} y_i \sqrt{2} \cos(j\pi t_i) - \mathrm{E}(\tilde{y}_j) \right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{r=1}^{n} \mathrm{E}(y_i - \mathrm{E}y_i)(y_r - \mathrm{E}y_r) \sqrt{2} \cos(j\pi t_i) \\
&= \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \sqrt{2} \cos(j\pi t_i) \\
&= 0, \quad j = 1, \ldots, n-1,
\end{aligned}$$

and

$$\mathrm{Cov}(\tilde{y}_j, \tilde{y}_k) = \mathrm{Cov}\left(\frac{1}{n} \sum_{i=1}^{n} y_i \sqrt{2} \cos(j\pi t_i), \frac{1}{n} \sum_{i=1}^{n} y_i \sqrt{2} \cos(k\pi t_i) \right)$$

$$
\begin{aligned}
&= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\sqrt{2}\cos(j\pi t_i) - \mathrm{E}(\tilde{y}_j)\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\sqrt{2}\cos(k\pi t_i) - \mathrm{E}(\tilde{y}_k)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{r=1}^{n} \mathrm{E}(y_i - \mathrm{E}y_i)(y_r - \mathrm{E}y_r)\sqrt{2}\cos(j\pi t_i)\sqrt{2}\cos(k\pi t_i) \\
&= \frac{\sigma^2}{n^2}\sum_{i=1}^{n} \sqrt{2}\cos(j\pi t_i)\sqrt{2}\cos(k\pi t_i) \\
&= \frac{\sigma^2}{n}\delta_{jk}, \quad j, k = 1, \ldots, n-1.
\end{aligned}
$$

Therefore we have

$$
\mathrm{IDR}(\lambda) \doteq \sum_{j=1}^{n-1} f_j^2 \left(\frac{2\lambda(j\pi)^2}{k_j^2 + 2\lambda(j\pi)^2}\right)^2 + \sum_{j=n}^{\infty}(f_j)^2 + \frac{\sigma^2}{n}\sum_{j=1}^{n-1}\frac{2k_j^2}{(k_j^2 + 2\lambda(j\pi)^2)^2}. \tag{2.32}
$$

and

$$
\mathrm{IRR}(\lambda) \doteq \sum_{j=1}^{n-1} \frac{1}{2}k_j^2 f_j^2 \left(\frac{2\lambda(j\pi)^2}{k_j^2 + 2\lambda(j\pi)^2}\right)^2 + \sum_{j=n}^{\infty}\frac{1}{2}k_j^2 f_j^2 + \frac{\sigma^2}{n}\sum_{j=1}^{n-1}\frac{k_j^4}{(k_j^2 + 2\lambda(j\pi)^2)^2}. \tag{2.33}
$$

To investigate the asymptotic behaviors of $\mathrm{IDR}(\lambda)$ and $\mathrm{IRR}(\lambda)$ we need to make some assumptions about the Fourier coefficients for f and K. Obviously $f_j^2 \to 0$ as $j \to \infty$, from Parseval's relation, since $\|f\|^2$ is bounded. However, integration by parts gives

$$
f_j = -\frac{1}{j\pi}f_j'
$$

with

$$
f_j' = \int_0^1 f'(t)\sqrt{2}\sin(j\pi t)\ dt, \quad j = 1, 2, \ldots,
$$

the Fourier coefficients for f' under the sine complete orthonormal sequences for $L_2[0,1]$. If we now model the f_j as having algebraic decay this entails that we assume

$$
f_j^2 \sim C_1 j^{-3-\alpha}, \quad j = 1, 2, \ldots \tag{2.34}
$$

for some positive constants C_1 and α. The assumption in (2.34) allows f_j^2 to decay algebraically in a manner that ensures $\|f'\|^2$ is bounded. For similar reason we also assume that

$$k_j^2 \sim C_2 j^{-3-\beta}, \quad j = 1, 2, \ldots \tag{2.35}$$

for some positive constants C_2 and β. Under condition (2.34),

$$\begin{aligned}
\sum_{j=n}^{\infty} f_j^2 &\sim \sum_{j=n}^{\infty} C_1 j^{-3-\alpha} \\
&\leq C_1 \int_{n-1}^{\infty} \frac{1}{x^{3+\alpha}} dx \\
&= C_1^*(n-1)^{-2-\alpha} \\
&\sim C_1^* n^{-2-\alpha},
\end{aligned}$$

for $C_1^* = -C_1/(2+\alpha)$. Similarly under conditions (2.34)-(2.35) we have

$$\begin{aligned}
\sum_{j=n}^{\infty} (\mu_j)^2 &= \sum_{j=n}^{\infty} (1/2)(k_j f_j)^2 \\
&\sim \sum_{j=n}^{\infty} (1/2) C_1 j^{-3-\alpha} C_2 j^{-3-\beta} \\
&\leq \frac{1}{2} C_1 C_2 \int_{n-1}^{\infty} \frac{1}{x^{6+\alpha+\beta}} dx \\
&= C_3 (n-1)^{-5-\alpha-\beta} \\
&\sim C_3 n^{-5-\alpha-\beta},
\end{aligned}$$

where $C_3 = -C_1 C_2/(2(5+\alpha+\beta))$. Applying (2.34)-(2.35) to the other terms in $\mathrm{IDR}(\lambda)$ and $\mathrm{IRR}(\lambda)$ as well, we obtain

$$\begin{aligned}
\mathrm{IDR}(\lambda) &\asymp \sum_{j=1}^{n-1} C_1 j^{-3-\alpha} \left(\frac{(2\lambda\pi^2) j^2}{C_2 j^{-3-\beta} + (2\lambda\pi^2) j^2} \right)^2 + C_1^* n^{-2-\alpha} \\
&\quad + \frac{\sigma^2}{n} \sum_{j=1}^{n-1} \frac{2 C_2 j^{-3-\beta}}{(C_2 j^{-3-\beta} + (2\lambda\pi^2) j^2)^2} \\
&= C_1 C_2^{-2} (2\pi^2)^2 \lambda^2 \sum_{j=1}^{n-1} \frac{j^{7-\alpha+2\beta}}{(1 + C_2^{-1}(2\lambda\pi^2) j^{5+\beta})^2} + C_1^* n^{-2-\alpha}
\end{aligned}$$

$$+ 2C_2^{-1}\frac{\sigma^2}{n}\sum_{j=1}^{n-1}\frac{j^{3+\beta}}{(1+C_2^{-1}(2\lambda\pi^2)j^{5+\beta})^2}$$

$$= D_1\lambda^2\sum_{j=1}^{n-1}\frac{j^{7-\alpha+2\beta}}{(1+uj^{5+\beta})^2} + C_1^* n^{-2-\alpha} + D_2 n^{-1}\sum_{j=1}^{n-1}\frac{j^{3+\beta}}{(1+uj^{5+\beta})^2}, \qquad (2.36)$$

where $u = C_2^{-1}(2\pi^2\lambda)$, $D_1 = C_1 C_2^{-2}(2\pi^2)^2$ and $D_2 = 2C_2^{-1}\sigma^2$. Similarly we have

$$\mathrm{IRR}(\lambda) \asymp \sum_{j=1}^{n-1}\frac{1}{2}C_1 j^{-3-\beta}C_2 j^{-3-\alpha}\left(\frac{(2\lambda\pi^2)j^2}{C_2 j^{-3-\beta}+(2\lambda\pi^2)j^2}\right)^2 + C_3 n^{-5-\alpha-\beta}$$

$$+ \frac{\sigma^2}{n}\sum_{j=1}^{n-1}\frac{C_2^2 j^{-6-2\beta}}{(C_2 j^{-3-\beta}+(2\lambda\pi^2)j^2)^2}$$

$$= \frac{C_1}{2C_2}(2\pi^2)^2\lambda^2\sum_{j=1}^{n-1}\frac{j^{4-\alpha+\beta}}{(1+C_2^{-1}(2\lambda\pi^2)j^{5+\beta})^2} + C_3 n^{-5-\alpha-\beta}$$

$$+ \frac{\sigma^2}{n}\sum_{j=1}^{n-1}\frac{1}{(1+C_2^{-1}(2\lambda\pi^2)j^{5+\beta})^2}$$

$$= D_3\lambda^2\sum_{j=1}^{n-1}\frac{j^{4-\alpha+\beta}}{(1+uj^{5+\beta})^2} + C_3 n^{-5-\alpha-\beta} + D_4 n^{-1}\sum_{j=1}^{n-1}\frac{1}{(1+uj^{5+\beta})^2}, \qquad (2.37)$$

where $D_3 = 2^{-1}C_1 C_2^{-1}(2\pi^2)^2$ and $D_4 = \sigma^2$.

We now need to deal with summations of the form $\sum_{k=1}^{\infty} k^s(1+uk^r)^{-2}$. For this purpose we can use the following lemma from Cox (1988).

Lemma 2.1. *Let*

$$D(u,s) = \sum_{k=1}^{\infty} k^s(1+uk^r)^{-2}.$$

Then $D(u,s) < \infty$ if $s < 2r-1$ for all $u > 0$. Furthermore, if $u \to 0^+$ then

$$D(u,s) \asymp \begin{cases} u^{-(s+1)/r} & \text{if } -1 < s < 2r-1, \\ log(1/u) & \text{if } s = -1, \\ 1 & \text{if } s < -1. \end{cases} \qquad (2.38)$$

Applying Lemma 2.1 to our expressions IDR(λ) in (2.36) and IRR(λ) in (2.37) produces the following results relating the optimal smoothing parameters for estimating f and μ.

Theorem 2.2. *Let* λ_f *and* λ_μ *be the optimal regularization parameters minimizing IDR*(λ) *and IRR*(λ)*, respectively. For* $\alpha > 0$ *and* $\beta > 0$*,*

(1) When $0 < \alpha < \beta + 5$*,*

$$\lambda_\mu \asymp \lambda_f = O(n^{-\frac{\beta+5}{\alpha+\beta+6}}), \tag{2.39}$$

(2) When $\alpha \geq \beta + 5$*,*

$$\lambda_\mu = o(\lambda_f). \tag{2.40}$$

The primary result of the Theorem is to provide a characterization of when the optimal smoothing parameter for f and μ have the same rates of decay. Theorem 2.2 states that this occurs when the decay exponent for f is in $(0, \beta+5)$. In this case one would expect to obtain similar asymptotic behavior by using, for example, a GCV estimators of λ for estimation of both μ and f.

In terms of the assumptions required for Theorem 2.2, larger values of α and β correspond to increased smoothness for f and K. Thus, when $\alpha < \beta+5$ this indicate that f has a high frequency composition in comparison to the kernel function K. Under these conditions, λ_μ also minimizes IDR(λ).

However, if $\beta \leq \alpha - 5$, either K has some high frequencies relative to f or f is too smooth compared to the smoothness of K. In other words, the frequencies of μ are mostly caused by the frequencies of K. Consequently λ_μ does not depend on α or the smoothness of f and $\lambda_\mu = o(\lambda_f)$. In this case estimation of f using λ_μ may not be effective. An example of result (3) can be seen by taking f to be a constant function, $f(\cdot) \doteq C$. In this case $\alpha = \infty$ and $f(\cdot)$ is very smooth or low frequency. The convolution of K and f behaves like CK having smoothness similar to that of K. As a result, λ_μ is not related to the smoothness of f.

Finally, using the fact that when $a + 1 > 0$ and $2 - (a+1) > 0$

$$\int_0^\infty \frac{y^a}{(1+y)^2}\, dy \;= \Gamma(a+1)\Gamma(2-(a+1)) \tag{2.41}$$

in Beyer (1991), we present a Theorem relating λ_μ and λ_f without using Lemma 2.1.

Theorem 2.3. *For λ_f and λ_μ as in Theorem 2.2 and $\alpha < \beta + 5$, $\lambda_\mu \asymp \lambda_f = O(n^{-\frac{\beta+5}{\alpha+\beta+6}})$.*

Let $\Delta = \{(\alpha, \beta) | \alpha < \beta + 5\}$. We showed that for any (α, β) in Δ, $\lambda_\mu \asymp \lambda_f = O(n^{-\frac{\beta+5}{\alpha+\beta+6}})$ in the proof of Theorem 2.3. We observe that any (α, β) satisfying conditions (1) in Theorem 2.2 belongs to Δ and $\lambda_\mu \asymp \lambda_f = O(n^{-\frac{\beta+5}{\alpha+\beta+6}})$. Thus, the results from Theorem 2.2 and Theorem 2.3 coincide when (α, β) is in Δ and satisfies conditions (1) in Theorem 2.2; that is when $(\alpha, \beta) \in \{(\alpha, \beta) | \alpha > 0, \beta > 0, \alpha < \beta + 5\}$. However when $\alpha \geq \beta + 5$ Theorem 2.2 applies while Theorem 2.3 does not. On the other hand, Theorem 2.3 applies for any $(\alpha, \beta) \in \Delta$ while Theorem 2.2 applies only in subset of Δ where α and β are positive, $\{(\alpha, \beta) | \alpha > 0, \beta > 0, \alpha < \beta + 5\}$.

2.2 Proofs

2.2.1 Proof of Proposition 2.1

Proof. If $u \in [0, 1]$ then

$$K(u) = k_0 + \sum_{j=1}^{\infty} k_j \sqrt{2} \cos(j\pi u)$$

since $K \in C^1[0, 1]$. For $u \in [-1, 0]$ we have

$$\begin{aligned} k_0 + \sqrt{2} \sum_{j=1}^{\infty} k_j \cos(j\pi u) &= k_0 + \sqrt{2} \sum_{j=1}^{\infty} k_j \cos(j\pi(-u)) \\ &= K(-u) \\ &= K(u) \end{aligned}$$

because K is symmetric about 0. If $u \in [1, 2]$ then

$$K(u) = K(2 - u)$$

and

$$k_0 + \sqrt{2}\sum_{j=1}^{\infty} k_j \cos(j\pi(2-u)) = k_0 + \sqrt{2}\sum_{j=1}^{\infty} k_j \cos(j\pi u).$$

$\square$

2.2.2 Proof of Theorem 2.1

Proof. To prove Theorem 2.1 we show

$$L_t(1) = k_0, \tag{2.42}$$

and

$$L_t\sqrt{2}\cos(j\pi(\cdot)) = k_j \cos(j\pi t), \quad j = 1, 2, \ldots \tag{2.43}$$

From the definition of the operator L_t in (2.1) and Proposition 2.1, we have

$$\begin{aligned}
L_t(1) &= \int_0^1 \frac{1}{2}\big(K(t-s) + K(t+s)\big)\ ds \\
&= \int_0^1 \frac{1}{2}\big(k_0 + \sum_{i=1}^{\infty} k_i\sqrt{2}\cos(i\pi(t-s)) + k_0 + \sum_{i=1}^{\infty} k_i\sqrt{2}\cos(i\pi(t+s))\big)\ ds \\
&= \int_0^1 \frac{1}{2}\big(k_0 + \sum_{i=1}^{\infty} k_i\sqrt{2}\cos(i\pi t)\cos(i\pi s) + \sum_{i=1}^{\infty} k_i\sqrt{2}\sin(i\pi t)\sin(i\pi s) \\
&\qquad + k_0 + \sum_{i=1}^{\infty} k_i\sqrt{2}\cos(i\pi t)\cos(i\pi s) - \sum_{i=1}^{\infty} k_i\sqrt{2}\sin(i\pi t)\sin(i\pi s)\big)\ ds \\
&= \int_0^1 \big(k_0 + \sum_{i=1}^{\infty} k_i\sqrt{2}\cos(i\pi t)\cos(i\pi s)\big)\ ds \\
&= k_0 + \sum_{i=1}^{\infty} k_i\sqrt{2}\cos(i\pi t)\int_0^1 \cos(i\pi s)\ ds \\
&= k_0,
\end{aligned}$$

since

$$\int_0^1 \cos(i\pi s)\ ds = 0.$$

Similarly,

$$L_t\sqrt{2}\cos(j\pi(\cdot)) = \int_0^1 \frac{1}{2}\big(K(t-s) + K(t+s)\big)\sqrt{2}\cos(j\pi s)\ ds$$

$$
\begin{aligned}
&= \int_0^1 \big(k_0 + \sum_{i=1}^{\infty} k_i \sqrt{2} \cos(i\pi t)\cos(i\pi s)\big) \sqrt{2} \cos(j\pi s)\ ds \\
&= k_0 \sqrt{2} \int_0^1 \cos(j\pi s)\ ds + 2 \sum_{i=1}^{\infty} k_i \cos(i\pi t) \int_0^1 \cos(i\pi s)\cos(j\pi s)\ ds \\
&= k_j \cos(j\pi t), \quad j = 1, 2, \ldots,
\end{aligned}
$$

because

$$
\int_0^1 \cos(i\pi s)\cos(j\pi s)\ ds = \frac{1}{2}\delta_{ij},
$$

where

$$
\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}
$$

Therefore, we have

$$
\begin{aligned}
L_t f &= L_t \big[f_0 + \sum_{j=1}^{\infty} f_j \sqrt{2} \cos(j\pi t)\big] \\
&= k_0 f_0 + \sum_{j=1}^{\infty} \frac{k_j f_j}{\sqrt{2}} \sqrt{2} \cos(j\pi t) \\
&= \mu_0 + \sum_{j=1}^{\infty} \mu_j \sqrt{2} \cos(j\pi t).
\end{aligned}
$$

The uniqueness of the Fourier series for $L_t f \in L_2[0,1]$ completes the proof. □

2.2.3 Proof of Corollary 2.1

Proof. We have

$$
\begin{aligned}
\mu_\lambda(t) &= \tilde{y}_0 + \sum_{j=1}^{n-1} \frac{k_j^2 \tilde{y}_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2} \cos(j\pi t) \\
&= \frac{1}{n} \sum_{i=1}^{n} y_i \big(1 + \sum_{j=1}^{n-1} \frac{k_j^2}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2} \cos(j\pi t) \sqrt{2} \cos(j\pi t_i)\big).
\end{aligned}
$$

The weight function for μ is

$$w_\mu(t,t_i) = 1 + \sum_{j=1}^{n-1} \frac{k_j^2}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2}\cos(j\pi t)\sqrt{2}\cos(j\pi t_i).$$

Similarly, we obtain

$$\begin{aligned} f_\lambda(t) &= \frac{\tilde{y}_0}{k_0} + \sum_{j=1}^{n-1} \frac{\sqrt{2}k_j\tilde{y}_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2}\cos(j\pi t) \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i\Big(\frac{1}{k_0} + \sum_{j=1}^{n-1} \frac{\sqrt{2}k_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2}\cos(j\pi t)\sqrt{2}\cos(j\pi t_i)\Big), \end{aligned}$$

and, hence,

$$w_f(t,t_i) = \frac{1}{k_0} + \sum_{j=1}^{n-1} \frac{\sqrt{2}k_j}{k_j^2 + 2\lambda(j\pi)^2} \sqrt{2}\cos(j\pi t)\sqrt{2}\cos(j\pi t_i).$$

□

2.2.4 Proof of Theorem 2.2

Proof. First we consider IDR(λ) when $\alpha/2 - \beta < 4$. We have

$$\begin{aligned} \mathrm{IDR}(\lambda) &\asymp D_1\lambda^2(C_2^{-1}2\pi^2\lambda)^{-\frac{2\beta-\alpha+8}{\beta+5}} + D_2 n^{-1}\lambda^{-\frac{\beta+4}{\beta+5}} + C_1^* n^{-2-\alpha} \\ &= C_1 C_2^{-\frac{\alpha+2}{\beta+5}}(2\pi^2)^{\frac{\alpha+2}{\beta+5}}\lambda^{\frac{\alpha+2}{\beta+5}} + 2\sigma^2 C_2^{-\frac{1}{\beta+5}}(2\pi^2)^{-\frac{1}{\beta+5}} n^{-1}\lambda^{-\frac{1}{\beta+5}} + C_1^* n^{-2-\alpha} \\ &= D_1^*\lambda^{\frac{\alpha+2}{\beta+5}} + D_2^* n^{-1}\lambda^{-\frac{\beta+4}{\beta+5}} + D^*, \end{aligned}$$

where $D_1^* = C_1 C_2^{-\frac{\alpha+2}{\beta+5}}(2\pi^2)^{\frac{\alpha+2}{\beta+5}}$, $D_2^* = 2\sigma^2 C_2^{-\frac{1}{\beta+5}}(2\pi^2)^{-\frac{1}{\beta+5}}$ and $D^* = C_1^* n^{-2-\alpha}$.

Thus,

$$\partial\frac{\mathrm{IDR}(\lambda)}{\partial\lambda} \asymp \frac{\alpha+2}{\beta+5} D_1^*\lambda^{-\frac{\beta-\alpha+3}{\beta+5}} - \frac{\beta+4}{\beta+5} D_2^* n^{-1}\lambda^{-\frac{2\beta+9}{\beta+5}}. \tag{2.44}$$

and, hence,

$$\lambda_f \asymp n^{-\frac{\beta+5}{\alpha+\beta+6}}.$$

Similarly if $\alpha/2 - \beta = 4$

$$\mathrm{IDR}(\lambda) \asymp D_1\lambda^2\log(C_2(2\pi^2)^{-1}\lambda^{-1}) + D_2^* n^{-1}\lambda^{-\frac{\beta+4}{\beta+5}} + D^*,$$

and

$$\partial\frac{\text{IDR}(\lambda)}{\partial\lambda} \asymp \lambda(2D_1\log(C_2(2\pi^2)^{-1}) + 2D_1\log\lambda^{-1} - D_1C_2^{-1}(2\pi^2)) \quad (2.45)$$
$$- \frac{\beta+4}{\beta+5}D_2^*n^{-1}\lambda^{-\frac{2\beta+9}{\beta+5}}$$
$$= \lambda(C_4 + C_5\log\lambda^{-1}) - C_6n^{-1}\lambda^{-\frac{2\beta+9}{\beta+5}}, \quad (2.46)$$

where $C_4 = C_1C_2^{-1}(2\pi^2)(2\log(C_2(2\pi^2)^{-1}) - C_2^{-1}(2\pi^2))$, $C_5 = 2C_1C_2^{-2}(2\pi^2)^2$ and $C_6 = \frac{\beta+4}{\beta+5}2\sigma^2C_2^{-\frac{1}{\beta+5}}(2\pi^2)^{-\frac{1}{\beta+5}}$. Therefore,

$$\lambda^{\frac{3\beta+14}{\beta+5}}C_6^{-1}(C_4 + C_5\log(\lambda^{-1})) \asymp n^{-1}.$$

In fact, we may observe that

$$\frac{\lambda^{\frac{3\beta+14}{\beta+5}}C_6^{-1}(C_4 + C_5\log(\lambda^{-1}))}{n^{-1}} \to 0, \text{ as } \lambda \to 0.$$

Consequently, in this case

$$\lambda_f = o(n^{-\frac{\beta+5}{3\beta+14}}).$$

For $\alpha/2 - \beta > 4$,

$$\text{IDR}(\lambda) \asymp D_1\lambda^2 + D_2^*n^{-1}\lambda^{-\frac{\beta+4}{\beta+5}} + D^*,$$

and

$$\partial\frac{\text{IDR}(\lambda)}{\partial\lambda} \asymp 2C_1C_2^{-2}(2\pi^2)^2\lambda - \frac{\beta+4}{\beta+5}D_2^*n^{-1}\lambda^{-\frac{2\beta+9}{\beta+5}}.$$

Thus,

$$\lambda_f \asymp n^{-\frac{\beta+5}{3\beta+14}}.$$

A similar analysis can be conducted for IRR(λ). When $\alpha - \beta < 5$, we have

$$\text{IRR}(\lambda) \asymp D_3\lambda^2(C_2^{-1}(2\pi^2)\lambda)^{-\frac{\beta-\alpha+5}{\beta+5}} + D_4n^{-1}\lambda^{-\frac{1}{\beta+5}} + C_3n^{-5-\alpha-\beta}$$
$$= D_3^*\lambda^{\frac{\beta+\alpha+5}{\beta+5}} + D_4n^{-1}\lambda^{-\frac{1}{\beta+5}} + D^{**},$$

where $D_3^* = D_3(C_2^{-1}(2\pi^2))^{-\frac{\beta-\alpha+5}{\beta+5}}$ and $D^{**} = C_3 n^{-5-\alpha-\beta}$,

$$\partial\frac{\mathrm{IRR}(\lambda)}{\partial\lambda} \asymp \frac{\alpha+\beta+5}{\beta+5}D_3^*\lambda^{\frac{\alpha}{\beta+5}} - \frac{1}{\beta+5}D_4 n^{-1}\lambda^{-\frac{\beta+6}{\beta+5}},$$

giving

$$\lambda_\mu \asymp n^{-\frac{\beta+5}{\alpha+\beta+6}}.$$

If $\alpha - \beta = 5$, we have

$$\mathrm{IRR}(\lambda) \asymp D_3\lambda^2\log(C_2(2\pi^2)^{-1}\lambda^{-1}) + D_4 n^{-1}\lambda^{-\frac{1}{\beta+5}} + D^{**},$$

$$\begin{aligned}\partial\frac{\mathrm{IRR}(\lambda)}{\partial\lambda} &\asymp \lambda(2D_3\log(C_2(2\pi^2)^{-1}) + 2D_3\log\lambda^{-1} - D_3C_2^{-1}(2\pi^2)) - \frac{1}{\beta+5}D_4 n^{-1}\lambda^{-\frac{\beta+6}{\beta+5}}\\ &= \lambda(C_7 + C_8\log\lambda^{-1}) - C_9 n^{-1}\lambda^{-\frac{\beta+6}{\beta+5}},\end{aligned}$$

where $C_7 = C_1C_2^{-1}(2\pi^2)^2(\log(C_2(2\pi^2)^{-1}) - 2^{-1}C_2^{-1}(2\pi^2))$, $C_8 = C_1C_2^{-1}(2\pi^2)^2$ and $C_9 = (\beta+5)^{-1}\sigma^2$. Thus,

$$\lambda^{\frac{2\beta+11}{\beta+5}}C_9^{-1}(C_7 + C_8\log\lambda^{-1}) \asymp n^{-1}$$

implies that

$$\lambda_\mu = o(n^{-\frac{\beta+5}{2\beta+11}}).$$

For $\alpha - \beta > 5$,

$$\mathrm{IRR}(\lambda) \asymp D_3\lambda^2 + D_4 n^{-1}\lambda^{-\frac{1}{\beta+5}} + D^{**},$$

$$\partial\frac{\mathrm{IRR}(\lambda)}{\partial\lambda} \asymp 2D_3\lambda - \frac{1}{\beta+5}D_4 n^{-1}\lambda^{-\frac{\beta+6}{\beta+5}},$$

and, hence,

$$\lambda_\mu \asymp n^{-\frac{\beta+5}{2\beta+11}}.$$

The relation between λ_f and λ_μ can be examined by studying their relative convergence rates as established above. For this purpose we examine the relations in several regions shown in Figure 1. When $\alpha \leq 5$, regardless of the range of β,

$$\lambda_\mu \asymp \lambda_f \asymp n^{-\frac{\beta+5}{\alpha+\beta+6}}.$$

The set of points (α, β) satisfying $\alpha - \beta < 5$ also satisfies $\alpha - 2\beta < 8$. In this region

$$\lambda_\mu \asymp \lambda_f \asymp n^{-\frac{\beta+5}{\alpha+\beta+6}}.$$

On the line $\beta = \alpha - 5$,

$$\lambda_f \asymp n^{-\frac{\beta+5}{2\beta+11}},$$

$$\lambda_\mu = o(n^{-\frac{\beta+5}{2\beta+11}}),$$

and therefore

$$\lambda_\mu = o(\lambda_f).$$

For $\alpha/2 - 4 < \beta < \alpha - 5$,

$$\lambda_\mu = o(\lambda_f)$$

since

$$\begin{aligned} \frac{\lambda_\mu}{\lambda_f} &\asymp \frac{n^{-\frac{\beta+5}{2\beta+11}}}{n^{-\frac{\beta+5}{\alpha+\beta+6}}} \\ &= \left(\frac{1}{n}\right)^{c^*(\beta+5)} \to 0 \text{ as } n \to \infty, \end{aligned}$$

where $c^* = \frac{\alpha-\beta-5}{(2\beta+11)(\alpha+\beta+6)} > 0$.

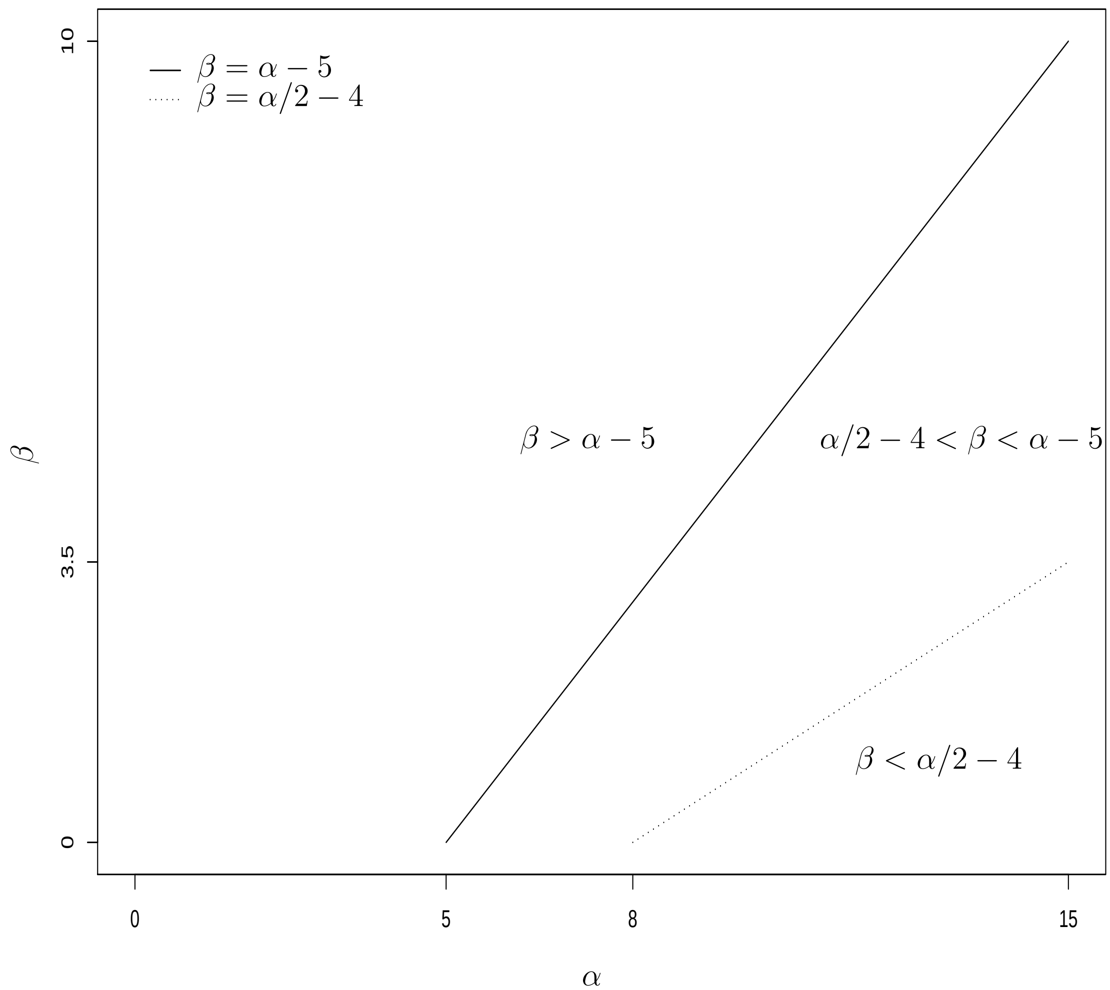

FIGURE 1: Selected regions of (α, β) for comparisons of relative convergence rates.

Similarly, on the line $\beta = \alpha/2 - 4$,

$$\lambda_f = o(n^{-\frac{\beta+5}{3\beta+14}}),$$

and

$$\lambda_\mu \asymp n^{-\frac{\beta+5}{2\beta+11}}.$$

Thus

$$\lambda_\mu = o(\lambda_f).$$

When $\beta < \alpha/2 - 4$,

$$\frac{\lambda_\mu}{\lambda_f} = \left(\frac{1}{n}\right)^{\frac{(\beta+3)(\beta+5)}{(2\beta+11)(3\beta+14)}} \to 0 \text{ as } n \to \infty,$$

and

$$\lambda_\mu = o(\lambda_f).$$

$\square$

2.2.5 Proof of Theorem 2.3

Proof. We conduct a very similar analysis as in 2.2.4. A change of variables and the fact stated in (2.41) give us

$$\begin{aligned}
\mathrm{IDR}(\lambda) &\asymp D_1\lambda^2 \sum_{j=1}^{n-1} \frac{j^{7-\alpha+2\beta}}{(1+uj^{5+\beta})^2} + C_1^* n^{-2-\alpha} + D_2 n^{-1} \sum_{j=1}^{n-1} \frac{j^{3+\beta}}{(1+uj^{5+\beta})^2} \\
&\asymp D_1\lambda^2 n \int_0^\infty \frac{x^{7-\alpha+2\beta}}{(1+ux^{5+\beta})^2} dx + C_1^* n^{-2-\alpha} + D_2 \int_0^\infty \frac{x^{3+\beta}}{(1+ux^{5+\beta})^2} dx \\
&= A_1 n \lambda^{\frac{\alpha+2}{\beta+5}} \int_0^\infty \frac{y^{\frac{\beta-\alpha+3}{\beta+5}}}{(1+y)^2} dy + C_1^* n^{-2-\alpha} + A_2 \lambda^{-\frac{(\beta+4)}{\beta+5}} \int_0^\infty \frac{y^{-\frac{1}{\beta+5}}}{(1+y)^2} dy,
\end{aligned}$$

where $A_1 = D_1(\beta+5)^{-1} C_2^{\frac{2\beta-\alpha+8}{\beta+5}} (2\pi^2)^{\frac{\alpha-2\beta-8}{\beta+5}}$ and $A_2 = D_2(\beta+5)^{-1} C_2^{\frac{\beta+4}{\beta+5}} (2\pi^2)^{-\frac{\beta+4}{\beta+5}}$. When $\beta > \alpha/2 - 4$, we have

$$\begin{aligned}
\mathrm{IDR}(\lambda) \asymp\ & A_1 n \lambda^{\frac{\alpha+2}{\beta+5}} \Gamma(\frac{2\beta-\alpha+8}{\beta+5}) \Gamma(\frac{\alpha+2}{\beta+5}) + C_1^* n^{-2-\alpha} \\
&+ A_2 \lambda^{-\frac{(\beta+4)}{\beta+5}} \Gamma(\frac{\beta+4}{\beta+5}) \Gamma(\frac{\beta+6}{\beta+5})
\end{aligned}$$

and

$$\begin{aligned}
\partial\frac{\mathrm{IDR}(\lambda)}{\partial\lambda} \asymp\ & A_1(\frac{\alpha+2}{\beta+5}) \Gamma(\frac{2\beta-\alpha+8}{\beta+5}) \Gamma(\frac{\alpha+2}{\beta+5}) n \lambda^{\frac{\alpha-\beta-3}{\beta+5}} \\
&- A_2(\frac{\beta+4}{\beta+5}) \Gamma(\frac{\beta+4}{\beta+5}) \Gamma(\frac{\beta+6}{\beta+5}) \lambda^{-\frac{2\beta+9}{\beta+5}}.
\end{aligned}$$

Thus,

$$\lambda_f \asymp n^{-\frac{\beta+5}{\alpha+\beta+6}}.$$

Similarly if $\beta > \alpha - 5$

$$\begin{aligned}
\mathrm{IRR}(\lambda) &\asymp D_3\lambda^2 \sum_{j=1}^{n-1} \frac{j^{4-\alpha+\beta}}{(1+uj^{5+\beta})^2} + C_3 n^{-5-\alpha-\beta} + D_4 n^{-1} \sum_{j=1}^{n-1} \frac{1}{(1+uj^{5+\beta})^2} \\
&\asymp D_3\lambda^2 n \int_0^\infty \frac{x^{4-\alpha+\beta}}{(1+ux^{5+\beta})^2} dx + C_3 n^{-5-\alpha-\beta} + D_4 \int_0^\infty \frac{1}{(1+ux^{5+\beta})^2} dx \\
&= A_3 n \lambda^{\frac{\alpha+\beta+5}{\beta+5}} \int_0^\infty \frac{y^{-\frac{\alpha}{\beta+5}}}{(1+y)^2} dy + C_3 n^{-5-\alpha-\beta} + A_4 \lambda^{-\frac{1}{\beta+5}} \int_0^\infty \frac{y^{-\frac{\beta+4}{\beta+5}}}{(1+y)^2} dy \\
&= A_3 n \lambda^{\frac{\alpha+\beta+5}{\beta+5}} \Gamma(\frac{\beta-\alpha+5}{\beta+5}) \Gamma(\frac{\beta+\alpha+5}{\beta+5}) + C_3 n^{-5-\alpha-\beta} \\
&\quad + A_4 \lambda^{-\frac{1}{\beta+5}} \Gamma(\frac{1}{\beta+5}) \Gamma(\frac{2\beta+9}{\beta+5}),
\end{aligned}$$

where $A_3 = D_3(\beta+5)^{-1} C_2^{\frac{\beta-\alpha+5}{\beta+5}} (2\pi^2)^{\frac{\alpha-\beta-5}{\beta+5}}$ and $A_4 = D_4(\beta+5)^{-1} C_2^{\frac{1}{\beta+5}} (2\pi^2)^{-\frac{1}{\beta+5}}$. Thus,

$$\begin{aligned}
\partial \frac{\mathrm{IRR}(\lambda)}{\partial \lambda} \asymp\; & A_3(\frac{\alpha+\beta+5}{\beta+5}) \Gamma(\frac{\beta-\alpha+5}{\beta+5}) \Gamma(\frac{\beta+\alpha+5}{\beta+5}) n \lambda^{\frac{\alpha}{\beta+5}} \\
& - A_4(\frac{1}{\beta+5}) \Gamma(\frac{1}{\beta+5}) \Gamma(\frac{2\beta+9}{\beta+5}) \lambda^{-\frac{\beta+6}{\beta+5}},
\end{aligned}$$

and, hence,

$$\lambda_\mu \asymp n^{-\frac{\beta+5}{\alpha+\beta+6}}.$$

□

CHAPTER III

COMPUTATIONS FOR APPROXIMATION OF METHOD OF REGULARIZATION ESTIMATORS

3.1 Introduction

Determination of nuclear magnetic resonance (NMR) relaxation of fluids in heterogeneous systems has applications in a wide variety of disciplines, including agriculture, petrophysics, colloid and surface science, food sciences, and biomedicine. For example, NMR relaxation measurements in biological tissues probe the nature of malignant changes (Inch, McCredie, Knispel, Thompson, and Pintar, 1974), the molecular dynamics of water in tissues (Gore, Brown, Zhong, and Armitage, 1989), and the structure of tissue (Belton and Ratcliffe, 1985). These measurements provide insight into the number and character of identifiable tissue compartments, or information about the dynamics of molecular exchange processes. Since nuclear spins in biological samples generally exist in many different environments, the measured magnetization gives rise to a spectrum of relaxation times. It is evident that a clear understanding of tissue spin relaxation requires reliable methods for data interpretation. A qualitative interpretation of the relaxation times in various tissues is utilized in diagnostic NMR imaging in medicine (Whittall, 1991), however, the quantitative and accurate interpretation of relaxation data could greatly increase the information that is gleaned from such studies.

In the past decade, the application of relaxation measurements to porous media has generated much interest (Watson and Chang, 1997). The characterization of pore structures is of fundamental importance for accurate evaluation of many chemical engineering processes, reservoir resources, and petroleum recovery, as well as for environmental remediation. One of the most important properties for characterizing a porous medium is the pore-size distribution. Pore-size distributions are very difficult to assess with conventional methods, such as mercury porosimetry, gas adsorption, or microscopic analysis of thin sections. Mercury porosimetry, which is based on a model of capillary tube bundles for data interpretation, tends to reflect pore throat sizes. Nitrogen adsorption can only detect a limited range of pore sizes and requires assumptions of pore shape; and microscopic analysis of thin sections does not allow for observation of the actual three-dimensional structures or facilitation of the sampling of reasonably large numbers of pores.

NMR relaxation measurement provides an effective and nondestructive means of probing a wide range of pore sizes (Gallegos and Smith, 1988; Banavar and Schwartz, 1989; Davies and Packer, 1990). Either longitudinal or transverse relaxation measurements can be used to determine pore-size distributions. Longitudinal relaxation has the advantage that it is unaffected by molecular diffusion in magnetic field gradients, while transverse relaxation is generally employed in downhole well-logging measurements because of the efficiency of the Carr-Purcell-Meiboom-Gill (CPMG) acquisition (Kleinberg, 1996). Pore-size distributions are estimated from relaxation data on the basis of an enhanced relaxation rate of pore fluid spins at the immediate vicinity of pore boundaries. In the fast-diffusion regime (Brownstein and Tarr, 1977), the enhanced relaxation rate exhibited by a single pore is directly related to the product of surface relaxivity and the pore surface-to-volume ratio. Porous media with a distribution of pore sizes have a nuclear magnetization relaxation that is nonexponential. A precise determination of the distribution of relaxation times is, therefore, critical for obtaining the information about pore-size distribution.

Estimation of the relaxation distribution from observed magnetization requires the solution of a Fredholm integral equation of the first kind. This is known to be an ill-posed problem in the sense that small changes in the observed data can result in large variations in the estimated distribution. Consequently, the methodology used for calculation of estimates from measured data can be critical.

To perform the estimation, a performance index is formulated based on, for example, least-squares or maximum likelihood estimation principles. The performance index reflects the precision of the match between measured data and the corresponding values calculated with a representation of the property. The estimate is chosen as the member of the solution space (i.e., the set of all admissible solutions) that optimizes (or minimizes) the performance index. It is desirable that the solution space be relatively large so that the true, although unknown, property lies within that space, or is accurately represented by a member of that space. Some applications use relatively simple functions, or a few discrete components represented as delta functions (Kleinberg, 1996; Timur, 1969). However, the actual number of components represented by the data is not known, and little confidence can be placed in the estimates if the property is not of the assumed form. Moreover, the selection of the location as well as strengths of the delta functions leads to a nonlinear optimization problem with some well-known accompanying difficulties regarding its solution. Other studies represent the candidate solutions with large numbers of delta functions, and impose regularization to stabilize the solution (Gallegos and Smith, 1988). With regularization, the performance index is augmented with a term proportional to the magnitude of the function estimates, or its derivatives. This has the advantage of providing for a larger solution space for the estimates. However, we expect the relaxation distributions in heterogeneous systems to be relatively smooth functions, a property not shared by estimates obtained by those methods. A better approach is to use B-splines for representing the solution space (Liaw et al., 1996). In this method, the estimates are continuous functions, rather than a set of discrete components. The selection of the regularization parameter, which determines the relative smoothing provided by the regularization term, is problematic. Various ad hoc and graphical methods have been proposed in the literature.

In this chapter, we present nonparametric regression techniques to develop a method for estimating relaxation distributions from NMR relaxation data. Throughout the chapter, theories and methodologies are not limited to the estimation of relaxation distributions but are developed to provide solutions for general, ill-posed inverse problems.

3.2 Approximation of MOR Estimators

3.2.1 Relaxation Distributions from NMR experiments

NMR longitudinal and transverse relaxation measurements are usually made using standard pulse sequences, i.e., inversion-recovery and CPMG sequences, respectively. The general expression for the magnetization μ at relaxation delay time t is given by

$$\mu(t) = \mu_0 \int_{\tau_{\min}}^{\tau_{\max}} f(\tau) K(t, \tau)\, d\tau, \tag{3.1}$$

where $f(\tau)$ is the unknown relaxation distribution, μ_0 is the limit of magnetization, $K(t, \tau)$ is the specified kernel function (Liaw et al., 1996). The experiment is called a T_1 or T_2 experiment when the relaxation delay time τ is for inversion-recovery measurements or CPMG measurements, respectively. The limits on the integral in (3.1) are chosen to contain the values of τ expected for the sample being studied.

To simplify the model and notation without loss of generality, we assume $\mu_0 = 1$ throughout the chapter. The relaxation distribution $f(\tau)$ is assumed to satisfy

$$\int_{\tau_{\min}}^{\tau_{\max}} f(\tau)\, d\tau = 1$$

and

$$f(\tau) \geq 0 \tag{3.2}$$

for all $\tau \in [\tau_{min}, \tau_{max}]$. The form of the kernel function is

$$K(t,\tau) = 1 - \alpha \exp\left(-\frac{t}{\tau}\right), \tag{3.3}$$

where $\alpha \doteq 2$ for inversion-recovery measurements and

$$K(t,\tau) = \exp\left(-\frac{t}{\tau}\right) \tag{3.4}$$

for CPMG measurements. In the T_1 case, we can either use $\alpha = 2$ in (3.3) or include it in the estimation problem. We then have

$$\begin{aligned} 1 - \mu(t) &= \int_{\tau_{\min}}^{\tau_{\max}} \alpha f(\tau) \exp\left(-\frac{t}{\tau}\right)\, d\tau && (3.5)\\ &= \int_{\tau_{\min}}^{\tau_{\max}} g(\tau) \exp\left(-\frac{t}{\tau}\right)\, d\tau, && (3.6)\end{aligned}$$

where

$$g(\tau) = \alpha f(\tau).$$

In this case, we estimate $g(\tau)$ and obtain a normalized relaxation density estimator of $f(\tau)$ using the relation

$$f(\tau) = \frac{g(\tau)}{\int_{\tau_{\min}}^{\tau_{\max}} g(\tau)\, d\tau}.$$

Our objective is to estimate the relaxation distribution function from discrete noisy observations of $\mu(t)$. This amounts to a solution of the Fredholm integral equation stated in (3.1). In natural media, we expect the relaxation distribution to be well represented by a smooth, continuous function. However, the form of that function is not known a priori, and should be part of the estimation process. In terms of approximating the MOR estimator, the usual approach is to find an approximation of the MOR estimator in a finite dimensional function subspace. We employ B-splines for this purpose but intend to develop an algorithm of sufficient generality to allow for other choices. We approximate the unknown relaxation distribution with p coefficients and basis function by

$$f(\tau) \doteq \sum_{j=1}^{p} \beta_j x_{rj}(\tau),$$

where the $\{x_{rj}\}_{j=1,p}$ are the B-spline basis functions for splines of order r with knots at points $\tau_1, \ldots, \tau_m$ in $[\tau_{\min}, \tau_{\max}]$, $p = r + m$ and the β_j are their corresponding coefficients. We assume p is large enough that it play no roles in the smoothing process.

Splines are piecewise polynomials, the different polynomial segments of which have been joined together at knots to ensure continuity properties. A spline of order r has $r - 2$ continuous derivatives providing the smoothest possible piecewise polynomial. It is also known that every continuous function on a finite domain can be approximated arbitrarily well by splines with sufficient numbers of knots (Schumaker, 1990). Some other nice properties of splines can be found in de Boor (1978).

3.2.2 Computations of Approximate MOR Estimators without Constraints

Let $y_1, \ldots, y_n$ represent the observed magnetization data at relaxation delay times $t_1, \ldots, t_n$ and assume that

$$\begin{aligned} y_i &= \mu(t_i) + \varepsilon_i && (3.7)\\ &= L_i f + \varepsilon_i \\ &= \int_{\tau_{\min}}^{\tau_{\max}} f(\tau) K(t_i, \tau)\ d\tau + \varepsilon_i \\ &\doteq \int_{\tau_{\min}}^{\tau_{\max}} \sum_{j=1}^{p} \beta_j x_{rj}(\tau) K(t_i, \tau)\ d\tau + \varepsilon_i \\ &= \sum_{j=1}^{p} \beta_j \int_{\tau_{\min}}^{\tau_{\max}} x_{rj}(\tau) K(t_i, \tau)\ d\tau + \varepsilon_i \\ &= \sum_{j=1}^{p} \beta_j L_i x_{rj} + \varepsilon_i, \quad i = 1, \ldots, n, && (3.8) \end{aligned}$$

where $L_i,\ i = 1, \ldots, n$, are the linear functionals representing the integration operator based on a kernel function K, and the $\varepsilon_i,\ i = 1, \ldots, n$, are zero mean, uncorrelated random variables with common variance σ^2. Rewriting the model shown in (3.8) using vector-matrix notation, we have

$$\mathbf{y} \doteq \mathbf{L}\beta + \varepsilon, \tag{3.9}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T,\ \mathbf{L} = \{L_i x_{rj}\}_{i=1,\ldots,n,j=1,\ldots,p},\ \beta = (\beta_1, \ldots, \beta_p)^T$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$. For a given $\lambda > 0$, an approximate MOR estimator is provided by the minimizer $\mathbf{b}_\lambda = (b_{\lambda 1}, \ldots, b_{\lambda p})^T$ with respect to $\mathbf{b} = (b_1, \ldots, b_p)^T$ of

$$n^{-1}(\mathbf{y} - \mathbf{L}\mathbf{b})^T(\mathbf{y} - \mathbf{L}\mathbf{b}) + \lambda \mathbf{b}^T \Omega \mathbf{b}, \tag{3.10}$$

for $\Omega = \{\int_{\tau_{\min}}^{\tau_{\max}} x_{ri}^{(m)}(\tau) x_{rj}^{(m)}(\tau)\ d\tau\}_{i=1,\ldots,p,\ j=1,\ldots,p}$. Differentiating (3.10) with respect to $\mathbf{b}$, we have

$$(\mathbf{L}^T\mathbf{L} + n\lambda\Omega)\mathbf{b}_\lambda = \mathbf{L}^T\mathbf{y}, \tag{3.11}$$

or assuming $(\mathbf{L}^T\mathbf{L} + n\lambda\Omega)$ is invertable,

$$\mathbf{b}_\lambda = (\mathbf{L}^T\mathbf{L} + n\lambda\Omega)^{-1}\mathbf{L}^T\mathbf{y}. \tag{3.12}$$

Consequently the approximate MOR estimate of the relaxation distribution is

$$f_\lambda(\tau) = \sum_{j=1}^{p} b_{\lambda j} x_{rj}(\tau)$$

for $\tau \in [\tau_{\min}, \tau_{\max}]$ with $\mathbf{b}_\lambda = (b_{\lambda 1}, \ldots, b_{\lambda p})^T$.

Estimation of the mean magnetization $\mu(\cdot)$ is obtained from

$$\begin{aligned} \mu_\lambda &= (\mu_\lambda(t_1), \ldots, \mu_\lambda(t_n))^T \\ &= (L_1 f_\lambda, \ldots, L_n f_\lambda)^T \\ &= \mathbf{L}\mathbf{b}_\lambda \end{aligned}$$

$$= \mathbf{L}(\mathbf{L}^T\mathbf{L} + n\lambda\Omega)^{-1}\mathbf{L}^T\mathbf{y}. \tag{3.13}$$

Equation (3.13) gives the form of a hat matrix $\mathbf{H}(\lambda)$ that transforms $\mathbf{y}$ responses to fitted value with

$$\mathbf{H}(\lambda) = \mathbf{L}(\mathbf{L}^T\mathbf{L} + n\lambda\Omega)^{-1}\mathbf{L}^T. \tag{3.14}$$

It is possible to compute $\mathbf{b}_\lambda$ using ordinary least-squares methods as we will now demonstrate. This approach also allows us to characterize the properties of the resulting soultion.

By definition, the matrix Ω is symmetric and positive semidefinite. Therefore we can write

$$\begin{aligned}\boldsymbol{\Omega} &= \boldsymbol{\Gamma}^T\mathbf{D}\boldsymbol{\Gamma}\\ &= (\mathbf{D}^{1/2}\boldsymbol{\Gamma})^T(\mathbf{D}^{1/2}\boldsymbol{\Gamma}),\end{aligned}$$

where $\mathbf{D}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{\Omega}$, $\mathbf{D}^{1/2}$ is a diagonal matrix with the square roots of the eigenvalues of $\boldsymbol{\Omega}$, and $\boldsymbol{\Gamma}$ is an orthogonal matrix. The matrix $\mathbf{D}^{1/2}$ exists since $\boldsymbol{\Omega}$ has p nonnegative eigenvalues. Let

$$\mathbf{A}_{(n+p)\times p}(\lambda) = \begin{bmatrix} \mathbf{L}_{n\times p} \\ \sqrt{n\lambda}(\mathbf{D}^{1/2}\boldsymbol{\Gamma})_{p\times p} \end{bmatrix}, \tag{3.15}$$

and

$$\mathbf{Y}_{(n+p)\times 1} = \begin{bmatrix} \mathbf{y}_{n\times 1} \\ \mathbf{0}_{p\times 1} \end{bmatrix}. \tag{3.16}$$

Then, we have

$$\| \mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b} \|^2 \equiv (\mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b})^T(\mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b}) \tag{3.17}$$

$$= (\mathbf{y} - \mathbf{L}\mathbf{b})^T(\mathbf{y} - \mathbf{L}\mathbf{b}) + n\lambda\mathbf{b}^T\boldsymbol{\Omega}\mathbf{b}. \tag{3.18}$$

We establish the existence of a unique solution of minimum length for

$$\| \mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b} \|^2, \tag{3.19}$$

Theorem 3.1. *Let the rank of* $\mathbf{A}_{(n+p)\times p}$ *be* $r \leq p$ *and let* $\mathbf{A}$ *have QR decomposition*

$$\mathbf{A}(\lambda) = \mathbf{QRS}^T,$$

where $\mathbf{Q}_{(n+p)\times(n+p)}$ *and* $\mathbf{S}_{p\times p}$ *are orthogonal matrices,* $\mathbf{R}_{(n+p)\times p}$ *has the form*

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

and $\mathbf{R}_{11}$ *is a* $r \times r$ *full rank matrix. Then, the unique solution of minimum length for (3.19) is*

$$\mathbf{b}_\lambda = \mathbf{S}[(\mathbf{R}_{11}^{-1}\mathbf{v}_1)^T, \mathbf{0}^T]^T$$

with

$$\begin{aligned} \mathbf{v} &= (\mathbf{v}_1^T, \mathbf{v}_2^T)^T \\ &= \mathbf{Q}^T\mathbf{Y}. \end{aligned}$$

The unique solution for (3.11) exists only when $(\mathbf{L}^T\mathbf{L} + n\lambda\Omega)$ is full rank. If $(\mathbf{L}^T\mathbf{L} + n\lambda\Omega)$ is not full rank, Theorem 3.1 shows that there exist a solution of minimum length for the problem minimizing (3.10). The solution does not depend on the choice of orthogonal decomposition. In practice a QR factorization of $\mathbf{A}(\lambda)$ is obtained using the Householder transformation which produces a factorization where $\mathbf{R}_{11}$ is an upper-triangular matrix. Therefore, the final solution is easily obtained by backward substitution to (3.58); see (3.58) -(3.60) in the proof of Theorem 3.1 in Section 3.6.

3.2.3 Computation of Approximate MOR Estimators with Inequality Constraints

An MOR estimator can often be qualitatively improved by the addition of constraints. When a function we are estimating is from experimental data, it is particularly necessary to impose certain constraints to the estimation process in order to have physically valid estimators. In our estimation of the relaxation distribution, non-negativity constraints are placed in equation (3.2). For a collection of points $\tau_1, \ldots, \tau_q$ in $[\tau_{min}, \tau_{max}]$, we can consider constraints

$$f(\tau_i) \geq 0, \quad i = 1, \ldots, q, \tag{3.20}$$

or equivalently

$$\mathbf{X}\beta \geq \mathbf{0}, \tag{3.21}$$

where

$$\mathbf{X} = \{x_j^m(\tau_i)\}_{i=1,\ldots,q,\ j=1,\ldots,p}.$$

For constraints (3.20) and (3.21) to be consistent with (3.2), it is necessary to have a sufficiently fine grid of evaluation points $\tau_1, \ldots, \tau_q$. The other way is to add a strong sufficient condition for the non-negativity of $f(\tau)$. Since B-spline bases are nonnegative, the constraints

$$\beta \geq \mathbf{0} \tag{3.22}$$

can provide a solution in the sense that the solution is consistent with (3.2) for all $\tau \in [\tau_{min}, \tau_{max}]$.

In this dissertation, we develop methods for obtaining an approximate MOR estimator subject to general inequality constraints for β of the form

$$\mathbf{G}_{q\times p}\beta_{p\times 1} \geq \mathbf{h}_{q\times 1},$$

where $\mathbf{G}$ is a known matrix and

$$\mathbf{h} = (h_1, \ldots, h_q)^T$$

is a specified vector. The criterion in (3.10) then becomes minimization of

$$n^{-1}(\mathbf{y} - \mathbf{L}\mathbf{b})^T(\mathbf{y} - \mathbf{L}\mathbf{b}) + \lambda\mathbf{b}^T\mathbf{\Omega}\mathbf{b}, \tag{3.23}$$

subject to

$$\mathbf{G}\mathbf{b} \geq \mathbf{h},$$

or equivalently minimization of

$$\parallel \mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b} \parallel^2, \tag{3.24}$$

subject to

$$\mathbf{G}\mathbf{b} \geq \mathbf{h}, \tag{3.25}$$

where $\mathbf{A}(\lambda)$ and $\mathbf{Y}$ are defined as in (3.15) and (3.16), respectively.

To solve problem (3.24)-(3.25), we will employ computational algorithms for constrained least-squares problems described in Lawson and Hanson (1974). The key idea is to solve sequences of equality constrained minimization problems by adding and dropping constraints iteratively until necessary conditions for the existence of the solution of (3.24) are satisfied. The sequences of equality constrained problems are solved by via unconstrained minimization methods that eliminates the equality constraints.

The primary algorithm we will employ is for obtaining a minimum norm solution $\hat{\mathbf{u}}$ to minimization of $\mathbf{u}^T\mathbf{u}$ subject to $\mathbf{E}\mathbf{u} \geq \mathbf{e}$. This is accomplished via solution of an appropriate dual problem as described on page 165 of the Lawson and Hanson (1974) text. However, in order to apply their algorithm, we must first transform our minimization problem into one that can be addressed by their methods.

First let us write

$$\mathbf{A}(\lambda) = \mathbf{Q} \begin{bmatrix} \mathbf{R} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{S}^T, \tag{3.26}$$

where $\mathbf{Q}_{(n+p)\times(n+p)}$ and $\mathbf{S}_{p\times p}$ are orthogonal matrices and $\mathbf{R}_{r\times r}$ is a full rank upper triangular matrix. We may then define

$$\mathbf{u} = \mathbf{S}^T\mathbf{b},$$

$$\mathbf{g} = \mathbf{Q}^T\mathbf{Y}$$

and partition the vectors $\mathbf{u}$, $\mathbf{g}$ as

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \begin{matrix} \} & r \\ \} & p-r \end{matrix}, \quad \mathbf{g} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \begin{matrix} \} & r \\ \} & p-r \end{matrix}$$

to obtain

$$\begin{aligned}
(\mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b})^T(\mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b}) &= \left(\mathbf{Q}\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{S}^T\mathbf{b} - \mathbf{Y}\right)^T \left(\mathbf{Q}\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{S}^T\mathbf{b} - \mathbf{Y}\right) \\
&= \left(\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{S}^T\mathbf{b} - \mathbf{Q}^T\mathbf{Y}\right)^T \left(\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{S}^T\mathbf{b} - \mathbf{Q}^T\mathbf{Y}\right) \\
&= (\mathbf{R}\mathbf{u}_1 - \mathbf{g}_1)^T(\mathbf{R}\mathbf{u}_1 - \mathbf{g}_1) + \mathbf{g}_2^T\mathbf{g}_2 \\
&= \| \mathbf{z} \|^2 + \| \mathbf{g}_2 \|^2,
\end{aligned}$$

where $\mathbf{z} = \mathbf{R}\mathbf{u}_1 - \mathbf{g}_1$. Note that the matrix $\mathbf{R}$ is full rank and upper triangular, so $\mathbf{R}^{-1}$ exists.

Now define

$$\begin{aligned}
\tilde{\mathbf{G}} &\equiv \mathbf{G}\mathbf{S}\begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&= [\,\overbrace{\tilde{\mathbf{G}}_1}^{r} \quad \overbrace{\tilde{\mathbf{G}}_2}^{p-r}\,].
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{G}\mathbf{b} &= \mathbf{G}\mathbf{S}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \\
&= \mathbf{G}\mathbf{S}\begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}
\end{aligned}$$

$$= \tilde{\mathbf{G}} \begin{bmatrix} \mathbf{R}\mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$$

and the original inequality constraints

$$\mathbf{G}\mathbf{b} \geq \mathbf{h}$$

become

$$\tilde{\mathbf{G}}_\mathbf{1}\mathbf{z} + \tilde{\mathbf{G}}_\mathbf{1}\mathbf{g}_1 + \tilde{\mathbf{G}}_\mathbf{2}\mathbf{u}_2 \geq \mathbf{h}.$$

We therefore employ the Lawson and Hanson (1974) methods to find the minimizer $\hat{\mathbf{z}}$ of $\mathbf{z}^T\mathbf{z}$ subject to

$$\tilde{\mathbf{G}}_\mathbf{1}\mathbf{z} \geq \mathbf{h} - \tilde{\mathbf{G}}_\mathbf{1}\mathbf{g}_1 - \tilde{\mathbf{G}}_\mathbf{2}\mathbf{u}_2$$

and the minimizer of $\mathbf{u}_2^T\mathbf{u}_2$ subject to

$$\tilde{\mathbf{G}}_\mathbf{2}\mathbf{u}_2 \geq \mathbf{h} - \tilde{\mathbf{G}}_\mathbf{1}\hat{\mathbf{z}} - \tilde{\mathbf{G}}_\mathbf{1}\mathbf{g}_1$$

and then iterate back and forth to solution.

3.3 Development of Hat Matrices

In the previous sections we have discussed how to compute the solutions of our minimization problem with inequality constraints. In this section we develop hat matrices for this MOR estimators settings. The key to accomplishing this is the Kuhn-Tucker theorem (Kuhn and Tucker, 1951) that tells us the constrained solution $\mathbf{b}_\lambda$ will be such that the elements of $\mathbf{G}\mathbf{b}_\lambda = \{c_i\}_{i=1,..,q}$ may be partitioned into two sets $\mathcal{E}^*$ and $\bar{\mathcal{E}}^*$ with $c_i = h_i$, $i \in \mathcal{E}^*$ and $c_i > h_i$, $i \in \bar{\mathcal{E}}^*$. Consequently, the minimizer $\mathbf{b}_\lambda$ of the inequality constrained minimization problem

$$n^{-1}(\mathbf{y} - \mathbf{L}\mathbf{b})^T(\mathbf{y} - \mathbf{L}\mathbf{b}) + \lambda \mathbf{b}^T \boldsymbol{\Omega} \mathbf{b}, \tag{3.27}$$

subject to

$$\mathbf{G}\mathbf{b} \geq \mathbf{h}$$

is also the minimizer of

$$n^{-1}(\mathbf{y} - \mathbf{L}\mathbf{b})^T(\mathbf{y} - \mathbf{L}\mathbf{b}) + \lambda \mathbf{b}^T \mathbf{\Omega} \mathbf{b}, \tag{3.28}$$

subject to

$$\mathbf{R}\mathbf{b} = \mathbf{r},$$

where $\mathbf{R} = \mathbf{G}_{\mathcal{E}^*}$ and $\mathbf{r} = \mathbf{h}_{\mathcal{E}^*}$ are the rows of $\mathbf{G}$ and $\mathbf{h}$ that correspond to the active constraints with indices in $\mathcal{E}^*$. Using this we obtain the following result concerning the form of estimator.

Theorem 3.2. *There exists the unique solution* $\mathbf{b}_\lambda$ *that achieves the global minimum of (3.27) or equivalently (3.28) and the form is*

$$\mathbf{b}_\lambda = \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^- \tilde{\mathbf{L}}^T\mathbf{y} - \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^- [\mathbf{I} - \mathbf{P}_\mathbf{R}] \left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right] \mathbf{R}^-\mathbf{r} + \mathbf{R}^-\mathbf{r}, \tag{3.29}$$

where

$$\begin{aligned}
\mathbf{R}^- &= \mathbf{R}^T(\mathbf{R}\mathbf{R}^T)^{-1}, \\
\mathbf{P}_\mathbf{R} &= \mathbf{R}^-\mathbf{R}, \\
\tilde{\mathbf{L}} &= \mathbf{L}(\mathbf{I} - \mathbf{P}_\mathbf{R}), \\
\tilde{\mathbf{\Omega}} &= (\mathbf{I} - \mathbf{P}_\mathbf{R})\mathbf{\Omega}(\mathbf{I} - \mathbf{P}_\mathbf{R}),
\end{aligned}$$

$(\cdot)^-$ *is a generalized inverse of* $(\cdot)$, *and* $\mathbf{R}$ *is as stated in (3.28).*

One may deduce the form of the constrained smoothing spline hat matrix from Theorem 3.2. In particular when $\mathbf{g} = \mathbf{0}$ we see that

$$\mathbf{y}_\lambda = \mathbf{L}\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^- \tilde{\mathbf{L}}^T\mathbf{y}$$

and the corresponding hat matrix $\hat{\mathbf{H}}(\lambda)$ is defined by

$$\hat{\mathbf{H}}(\lambda) = \mathbf{L}\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^{-}\tilde{\mathbf{L}}^T, \tag{3.30}$$

where $\tilde{\mathbf{L}}$ and $\tilde{\boldsymbol{\Omega}}$ are as stated in Theorem 3.2.

More generally, the hat matrix can be developed for general constraints based on transformations. We define $\mathbf{z}$ as the transformation of $\mathbf{y}$ such that

$$\mathbf{z} = \left[\mathbf{L}^T\mathbf{y} - \mathbf{L}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - n\lambda\boldsymbol{\Omega}\mathbf{R}^{-}\mathbf{r}\right].$$

We have $(\mathbf{I} - \mathbf{P_R})^{-} = (\mathbf{I} - \mathbf{P_R})$ since the matrix $(\mathbf{I} - \mathbf{P_R})$ is symmetric and idempotent. We observe that

$$\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^{-}(\mathbf{I} - \mathbf{P_R})\mathbf{z} = \mathbf{b}_\lambda - \mathbf{R}^{-}\mathbf{r}$$

and

$$\begin{aligned}\mathbf{y}_\lambda &= \mathbf{L}\mathbf{b}_\lambda \\ &= \tilde{\mathbf{H}}(\lambda)\mathbf{z} + \mathbf{L}\mathbf{R}^{-}\mathbf{r}\end{aligned}$$

where

$$\tilde{\mathbf{H}}(\lambda) = \mathbf{L}\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^{-}(\mathbf{I} - \mathbf{P_R}).$$

Let $\mathbf{z}_\lambda$ be $\mathbf{y}_\lambda - \mathbf{L}\mathbf{R}^{-}\mathbf{r}$. Then we have

$$\mathbf{z}_\lambda = \tilde{\mathbf{H}}(\lambda)\mathbf{z}.$$

We summarize these results in the next Corollary.

Corollary 3.1. *There exists a hat matrix $\tilde{\mathbf{H}}(\lambda)$ that maps $\mathbf{z}$ to $\mathbf{z}_\lambda$, where $\mathbf{z}$ and $\mathbf{z}_\lambda$ are transformations of $\mathbf{y}$ and $\mathbf{y}_\lambda$, respectively, such that*

$$\mathbf{z} = \left[\mathbf{L}^T\mathbf{y} - \mathbf{L}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - n\lambda\boldsymbol{\Omega}\mathbf{R}^{-}\mathbf{r}\right],$$

$$\mathbf{z}_\lambda = \mathbf{y}_\lambda - \mathbf{L}\mathbf{R}^-\mathbf{r},$$

and the hat matrix $\tilde{\mathbf{H}}(\lambda)$ *is defined by*

$$\tilde{\mathbf{H}}(\lambda) = \mathbf{L}\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^- (\mathbf{I} - \mathbf{P_R}).$$

Corollary 3.1 gives a solution for a closed form of the hat matrix for transformed responses $\mathbf{z}$ and predictions of responses $\mathbf{z}_\lambda$. Note that $\mathbf{z}$ here can be regarded as representing the unique information in the data after information from the constraints is removed.

3.4 Selection of the Smoothing Parameter

We have shown how to calculate the approximate MOR estimator $\mathbf{b}_\lambda$ for a given smoothing parameter λ. We now address the issue of choosing λ adaptively from the data.

Data-driven criteria for the selection of λ play a very important role in the performance of the estimator. For example, if we use a very small value of λ in the minimization criterion in (3.10), then the result will generally be too rough. In such a case we are primarily seeking a solution that minimizes $n^{-1}\sum_{i=1}^n (y_i - L_i f_\lambda)^2$. As a result, the estimated function f_λ may lose its physical validity. On the other hand, if a very large λ is employed this may result in an estimate f_λ that is too smooth. We lose data information in this case since the minimization in (3.10) corresponds primarily to minimization of the regularization term.

There are several ways of selecting a value for the smoothing parameter. One might decide upon a value for λ after trying many values of λ in an ad-hoc fashion. The basis for the decision can be one's experience or the expected degree of smoothness of the function being estimated based on some physical background. However, this method would require a lot of time and thus would not be very effective, particularly since the value of λ can be any finite positive real number. Choosing one value from such a big range would not be an easy task.

In this chapter we develop a criterion for the choice of λ, especially when constrained smoothing splines are employed. To discuss the selection of a smoothing parameter further, we present the necessary motivation for developing our proposed criterion. First, we consider the unconstrained model for the approximate MOR estimator in (3.9), minimization criterion in (3.10) and the resulting hat matrix in (3.14). The responses $y_1, \ldots, y_n$ are observed values from an unknown function μ at design points $t_1, \ldots, t_n$ with random errors. Consider an estimator for f,

$$f_\lambda(\cdot) = \sum_{j=1}^{p} b_{\lambda j} x_{rj}(\cdot),$$

with

$$\mathcal{L} f_\lambda = (L_1 f_\lambda, \ldots, L_n f_\lambda)^T \tag{3.31}$$

$$= \mathbf{H}(\lambda)\mathbf{y}, \tag{3.32}$$

where $\lambda \in (0, \infty)$. Note that the $\mathcal{L}$ in (3.31) is introduced to simplify notation and that it is different from the $\mathbf{L}$ in (3.9). The next question is how to choose λ in $(0, \infty)$ to find a good estimator f_λ

An ideal criterion for choosing λ might be the minimizer of integrated domain loss

$$\mathrm{IDL}(\lambda) = \int_{\tau_{\min}}^{\tau_{\max}} [f(\tau) - f_\lambda(\tau)]^2 \, d\tau \tag{3.33}$$

or the integrated domain risk

$$\mathrm{IDR}(\lambda) = \int_{\tau_{\min}}^{\tau_{\max}} \mathrm{E}[f(\tau) - f_\lambda(\tau)]^2 \, d\tau. \tag{3.34}$$

Unfortunately, there is no known practical estimator of $\mathrm{IDL}(\lambda)$ or $\mathrm{IDR}(\lambda)$. One possible way might be to find a matrix $\mathbf{C}$ such that

$$\mathrm{E}\mathbf{C}\mathbf{y} = f.$$

However, this can be effectively accomplished only for a few well-understood, special cases and it in general may lead to very ill-posed problems.

Thus, instead of considering the distance between f and f_λ directly, we consider distance measures between $\mathcal{L}f$ and $\mathcal{L}f_\lambda$, where $\mathcal{L}$ is as in (3.31). First we can think about the squared Euclidean distance between $\mathcal{L}f$ and $\mathcal{L}f_\lambda$ or, equivalently, the loss defined by

$$\mathrm{L}(\lambda) = n^{-1}(\mathcal{L}f - \mathcal{L}f_\lambda)^T(\mathcal{L}f - \mathcal{L}f_\lambda) \tag{3.35}$$

$$= n^{-1}(\mu - \mathcal{L}f_\lambda)^T(\mu - \mathcal{L}f_\lambda). \tag{3.36}$$

Or we can consider the risk, i.e., the expected value of the loss defined by

$$\mathrm{R}(\lambda) = \mathrm{E}(\mathrm{L}(\lambda)). \tag{3.37}$$

A minimizer of $\mathrm{L}(\lambda)$ provides a good estimator for a particular data set and a minimizer of $\mathrm{R}(\lambda)$ provides a good estimator for the prediction of future responses or in repeated samplings.

When we want a good estimator as a predictor of a future set of observations the prediction risk can be considered. In this case, we are looking for a predictor of

$$\mathbf{y}^* = \mathcal{L}f + \varepsilon^*, \tag{3.38}$$

where ε^* is a vector of random errors with zero means that are uncorrelated with each other and with ε, and have common variance, σ^2. The prediction risk is defined by

$$\mathrm{P}(\lambda) = n^{-1}(\mathbf{y}^* - \mathcal{L}f_\lambda)^T(\mathbf{y}^* - \mathcal{L}f_\lambda). \tag{3.39}$$

The relation between $\mathrm{P}(\lambda)$ and $\mathrm{R}(\lambda)$ is that

$$\mathrm{P}(\lambda) = \mathrm{R}(\lambda) + \sigma^2,$$

and, therefore, the minimizer of $\mathrm{P}(\lambda)$ and $\mathrm{R}(\lambda)$ are the same.

In practice, L(λ), R(λ) and P(λ) cannot be used directly since the mean function $\mathcal{L}f$ is unknown. Under our models in (3.7) we see E$\mathbf{y} = \mu$. Thus, P(λ) might be estimated by the SSR(λ) where

$$\begin{aligned}\mathrm{SSR}(\lambda) &= (\mathbf{y} - \mathcal{L}f_\lambda)^T(\mathbf{y} - \mathcal{L}f_\lambda)\\ &= (\mathbf{y} - \mathbf{H}(\lambda)\mathbf{y})^T(\mathbf{y} - \mathbf{H}(\lambda)\mathbf{y})\\ &= \mathbf{y}^T(\mathrm{I} - \mathbf{H}(\lambda))^2\mathbf{y}.\end{aligned}$$

To access the performance of $\frac{1}{n}$SSR(λ) as an estimator of $P(\lambda)$ we observe that

$$\begin{aligned}\mathrm{E}(\mathrm{SSR}(\lambda)) &= \mu^T(\mathrm{I} - \mathbf{H}(\lambda))^2\mu + \sigma^2\mathrm{tr}[(\mathrm{I} - \mathbf{H}(\lambda))^2]\\ &= \mu^T(\mathrm{I} - \mathbf{H}(\lambda))^2\mu + n\sigma^2 + \sigma^2\mathrm{tr}[\mathbf{H}(\lambda)^2] - 2\sigma^2 tr[\mathbf{H}(\lambda)]\end{aligned}$$

and

$$\begin{aligned}\mathrm{P}(\lambda) &= \sigma^2 + \mathrm{R}(\lambda)\\ &= \sigma^2 + \frac{1}{n}\mathrm{E}(\mu - \mathcal{L}f_\lambda)^T(\mu - \mathcal{L}f_\lambda)\\ &= \sigma^2 + \frac{1}{n}\mu^T(\mathrm{I} - \mathbf{H}(\lambda))^2\mu + \frac{\sigma^2}{n}\mathrm{tr}[\mathbf{H}(\lambda)^2].\end{aligned}$$

Then an unbiased estimator of P(λ) is

$$\hat{\mathrm{P}}(\lambda) = \frac{1}{n}\mathrm{SSR}(\lambda) + \frac{2\sigma^2}{n}\mathrm{tr}[\mathbf{H}(\lambda)],$$

since the bias is $-\frac{2\sigma^2}{n}\mathrm{tr}[\mathbf{H}(\lambda)]$ and an unbiased estimator of the risk is

$$\hat{\mathrm{R}}(\lambda) = \frac{1}{n}\mathrm{RSS}(\lambda) + \frac{2\sigma^2}{n}\mathrm{tr}[H(\lambda)] - \sigma^2.$$

The generalized cross-validation (GCV) criterion was proposed by Craven and Wahba (1979). Assuming that tr[$\mathbf{H}(\lambda)$] $< n$, GCV(λ) is defined by

$$\mathrm{GCV}(\lambda) = \frac{(1/n)\sum_{i=1}^n [y_i - L_i f_\lambda]^2}{[1 - (1/n)\mathrm{tr}\mathbf{H}(\lambda)]^2}.$$

It may be shown that $\mathrm{GCV}(\lambda)$ is a (biased) estimator of $\mathrm{P}(\lambda)$ that has various consistency and optimality properties as demonstrated, e.g., by Craven and Wahba (1979), Golub, Heath, and Wahba (1979), Nychka (1984) and Rice (1984).

To use $\hat{\mathrm{P}}(\lambda)$ we will generally require an estimate of σ^2. For this purpose one may use the Gasser, Sroka, and Jennen-Steinmetz (1986) estimator of σ^2 based on pseudo-residuals. Defining the i-th pseudo-residual $\tilde{\varepsilon}_i$ to be

$$\tilde{\varepsilon}_i = \mathrm{c}_i y_{i-1} + \mathrm{d}_i y_{i+1} - y_i, \quad i = 1, \ldots, n-1,$$

where

$$\mathrm{c}_i = \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}}$$

and

$$\mathrm{d}_i = \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}$$

we obtain the variance estimator

$$\hat{\sigma}^2 = (n-2)^{-1} \sum_{i=2}^{n-1} \frac{\tilde{\varepsilon}_i^2}{\mathrm{c}_i^2 + \mathrm{d}_i^2 + 1}.$$

Gasser et al. (1986) found that under mild restrictions, $\hat{\sigma}^2$ is a $\sqrt{n}$-consistent estimator of σ^2 and $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, V)$.

We now extend the previous data-driven criteria to include MOR estimators from (3.27) or equivalently from (3.28). We consider the same transformations of $\mathbf{y}$ and $\mathbf{y}_\lambda$ that depend on $\mathbf{R}^-\mathbf{r}$ in Corollary 3.1. More specifically,

$$\mathbf{z} = \left[\mathbf{L}^T\mathbf{y} - \mathbf{L}^T\mathbf{L}\mathbf{R}^-\mathbf{r} - n\lambda\mathbf{\Omega}\mathbf{R}^-\mathbf{r}\right], \tag{3.40}$$

$$\mathbf{z}_\lambda = \mathbf{y}_\lambda - \mathbf{L}\mathbf{R}^-\mathbf{r} \tag{3.41}$$

and the hat matrix $\tilde{\mathbf{H}}(\lambda)$ is defined by

$$\tilde{\mathbf{H}}(\lambda) = \mathbf{L}\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^- (\mathbf{I} - \mathbf{P}_{\mathbf{R}}). \tag{3.42}$$

Since minimizing $(\mathbf{y}-\mathbf{H}(\lambda)\mathbf{y})^T(\mathbf{y}-\mathbf{H}(\lambda)\mathbf{y})$ subject to $\mathbf{R}\beta=\mathbf{r}$ is the same as minimizing $(\mathbf{z}-\tilde{\mathbf{H}}(\lambda)\mathbf{z})^T(\mathbf{z}-\tilde{\mathbf{H}}(\lambda)\mathbf{z})$, this suggests the following approach to selecting λ for constrained problems.

Let $\mathbf{z}$, $\mathbf{z}_\lambda$ and $\tilde{\mathbf{H}}(\lambda)$ be as in (3.40), (3.41) and (3.42), respectively. With the active constraint

$$\mathbf{R}\mathbf{b}=\mathbf{r}, \tag{3.43}$$

resulting from an inequality constrained problem. Then we define

$$\hat{\mathrm{P}}_C(\lambda)=n^{-1}(\mathbf{z}-\mathbf{z}_\lambda)^T(\mathbf{z}-\mathbf{z}_\lambda)+\frac{2\sigma^2}{n}\mathrm{tr}[\tilde{\mathbf{H}}(\lambda)],$$

and

$$\mathrm{GCV}_C(\lambda)=\frac{(1/n)\sum_{i=1}^n\left[z_i-z_{\lambda i}\right]^2}{\left[1-(1/n)\mathrm{tr}\tilde{\mathbf{H}}(\lambda)\right]^2} \tag{3.44}$$

as parallels of $\hat{\mathrm{P}}(\lambda)$ and $\mathrm{GCV}(\lambda)$ that can be used in the constrained setting.

The use of these criteria would appear to entail calculations involving generalized inverses and the trace of $\tilde{\mathbf{H}}(\lambda)$. In the special, but important case when $\mathbf{r}=\mathbf{0}$, there are simplifications and we find that

$$\begin{aligned}\hat{\mathbf{H}}(\lambda)&=\mathbf{L}(\tilde{\mathbf{L}}^T\tilde{\mathbf{L}}+n\lambda\tilde{\mathbf{\Omega}})^{-}\tilde{\mathbf{L}}^T\\&=\mathbf{L}(\mathbf{I}-\mathbf{P}_\mathbf{R})(\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega})^{-}(\mathbf{I}-\mathbf{P}_\mathbf{R})\mathbf{L}^T\\&=\mathbf{L}(\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega})^{-1}\mathbf{L}^T-\mathbf{L}\mathbf{P}_\mathbf{R}(\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega})^{-1}\mathbf{L}^T\\&\quad-\mathbf{L}(\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega})^{-1}\mathbf{P}_\mathbf{R}\mathbf{L}^T+\mathbf{L}\mathbf{P}_\mathbf{R}(\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega})^{-1}\mathbf{P}_\mathbf{R}\mathbf{L}^T.\end{aligned}$$

Therefore, we observe

$$\mathrm{tr}(\hat{\mathbf{H}}(\lambda))=\mathrm{tr}(\mathbf{H}(\lambda))-\mathrm{tr}\Big\{(\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega})^{-1}\mathbf{L}^T\mathbf{L}\mathbf{P}_\mathbf{R}\Big\}. \tag{3.45}$$

Consequently, the problem becomes one of computing the trace for the unconstrained estimator and MOR "fit" to the columns of $\mathbf{L}\mathbf{P}_\mathbf{R}$.

3.5 Bayesian Credible Intervals

As a tool for inference about estimated functions we present point-wise Bayesian credible intervals for the mean function and the associated inverse function similar to those of Wahba (1983). These Bayesian credible intervals can be used to reflect the variability and uncertainty of our function estimators.

First we develop a Bayesian model that is related to our MOR estimator. For a normally distributed $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\mathbf{y}$ is distributed n-variate normal $\mathbf{N}_n(\mathbf{L}\beta, \sigma^2\mathbf{I})$. The likelihood function of $\mathbf{y}$ is

$$p_\beta(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{L}\beta)^T(\mathbf{y} - \mathbf{L}\beta)\right].$$

To motivate the prior distribution of β, we see that the minimization of criterion (3.10) is the same as maximizing

$$\exp\left\{-\frac{1}{2\sigma^2}\left\{(\mathbf{y} - \mathbf{L}\beta)^T(\mathbf{y} - \mathbf{L}\beta) + n\lambda\beta^T\mathbf{\Omega}\beta\right\}\right\}. \tag{3.46}$$

Thus, up to proportionality constants, we set the prior for β to be

$$f(\beta) \propto \exp\left\{-\frac{n\lambda}{2\sigma^2}\beta^T\mathbf{\Omega}\beta\right\}.$$

The joint density for $\mathbf{y}$ and β is then proportional to

$$\begin{aligned}
&\exp\left\{-\frac{n\lambda}{2\sigma^2}\beta^T\mathbf{\Omega}\beta\right\}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{L}\beta)^T(\mathbf{y} - \mathbf{L}\beta)\right\}\\
&= \exp\left\{-\frac{n\lambda}{2\sigma^2}\beta^T\mathbf{\Omega}\beta\right\}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}^T\mathbf{y} - 2\beta^T\mathbf{L}^T y + \beta^T\mathbf{L}^T\mathbf{L}\beta)\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(n\lambda\beta^T\mathbf{\Omega}\beta - \mathbf{y}^T\mathbf{y} - 2\beta^T\mathbf{L}^T\mathbf{y} + \beta^T\mathbf{L}^T\mathbf{L}\beta\right)\right\}\\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(\beta^T\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]\beta - 2\beta^T\mathbf{L}^T\mathbf{y}\right)\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{y}^T\mathbf{y}\right)\right\}\\
&= \exp\left\{-\frac{1}{2}\left(\beta - \left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{L}^T\mathbf{y}\right)^T\frac{\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]}{\sigma^2}\left(\beta - \left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{L}^T\mathbf{y}\right)\right\}\\
&\quad\cdot\exp\left\{-\frac{1}{2\sigma^2}\left[\left(\mathbf{y}^T\mathbf{y}\right) - (\mathbf{L}^T\mathbf{y})^T\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}(\mathbf{L}^T\mathbf{y})\right]\right\}.
\end{aligned}$$

Therefore, we obtain the posterior distribution

$$f(\beta|\mathbf{y}) = N_p\left(\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{L}^T\mathbf{y}, \sigma^2\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}\right) \tag{3.47}$$

which gives the posterior mean

$$\begin{aligned}\mathrm{E}(\beta|\mathbf{y}) &= \beta_\lambda \\ &= \left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{L}^T\mathbf{y}\end{aligned}$$

as a Bayes estimator of β and a posterior variance-covariance matrix

$$\mathrm{Cov}(\beta|\mathbf{y}) = \sigma^2\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}.$$

The posterior mean is the maximum likelihood estimator of β since it is the mean of the p-variate normal distribution in (3.47).

Let $\mathbf{S}(\lambda)$ be the matrix $\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}$ and $s_{ii}(\lambda)$ be the i-th diagonal element of the $\mathbf{S}(\lambda)$ matrix. For a known σ^2, a $100(1-\alpha)\%$ Bayesian credible interval for the i-th element β_i of β is

$$\beta_{\lambda i} \pm z_{\alpha/2}\sigma\sqrt{s_{ii}(\lambda)},$$

where $z_{\alpha/2}$ denotes the $100(1-\alpha/2)\%$ quantile of the standard normal distribution. Furthermore, for a specified vector $\mathbf{a} = (a_1, \ldots, a_p)^T$, we have

$$\begin{aligned}\mathrm{E}(\mathbf{a}^T\beta|\mathbf{y}) &= \mathbf{a}^T\mathrm{E}(\beta|\mathbf{y}) \\ &= \mathbf{a}^T\beta_\lambda\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(\mathbf{a}^T\beta|\mathbf{y}) &= \mathbf{a}^T\mathrm{Cov}(\beta|\mathbf{y})\mathbf{a} \\ &= \sigma^2\mathbf{a}^T\left[\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{a}.\end{aligned}$$

Therefore, a $100(1-\alpha)\%$ Bayesian credible interval for $\mathbf{a}^T\beta$ is

$$\mathbf{a}^T\beta_\lambda \pm z_{\alpha/2}\sigma\sqrt{\mathbf{a}^T\left[\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{a}}.$$

In particular, if we let $\mathbf{l}_i^T$ be the i-th row vector of $\mathbf{L}$ so that

$$\mu(t_i) = \mathbf{l}_i^T\beta$$

and

$$\mu_\lambda(t_i) = \mathbf{l}_i^T\beta_\lambda, \tag{3.48}$$

a $100(1-\alpha)\%$ Bayesian credible interval for $\mu(t_i)$ is

$$\mu_\lambda(t_i) \pm z_{\alpha/2}\sigma\sqrt{\mathbf{l}_i^T\left[\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{l}_i} \tag{3.49}$$

or, equivalently,

$$\mu_\lambda(t_i) \pm z_{\alpha/2}\sigma\sqrt{h_{ii}(\lambda)}, \tag{3.50}$$

where $h_{ii}(\lambda)$ is the i-th diagonal element of the hat matrix $\mathbf{H}(\lambda) = \mathbf{L}\left[\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{L}^T$.

Now let the vector of basis functions for splines of order r at an evaluational point τ be

$$\mathbf{x}_r^T(\tau) = (x_{r1}(\tau), \ldots, x_{rp}(\tau)). \tag{3.51}$$

Then the approximate MOR estimator at τ is

$$f_\lambda(\tau) = \mathbf{x}^T(\tau)\beta_\lambda, \tag{3.52}$$

and a $100(1-\alpha)\%$ Bayesian credible interval for $f(\tau)$ is

$$f_\lambda(\tau) \pm z_{\alpha/2}\sigma\sqrt{\mathbf{x}^T(\tau)\left[\mathbf{L}^T\mathbf{L}+n\lambda\mathbf{\Omega}\right]^{-1}\mathbf{x}(\tau)}. \tag{3.53}$$

We can also develop Bayesian credible intervals for a constrained model using techniques similar to those of the proof of Theorem 3.2. The minimization criterion

$$(\mathbf{y}-\mathbf{L}\mathbf{b})^T(\mathbf{y}-\mathbf{L}\mathbf{b}) + n\lambda\mathbf{b}^T\mathbf{\Omega}\mathbf{b}$$

subject to

$$\mathbf{R}\mathbf{b} = \mathbf{r},$$

is the same as

$$\max_{\beta} \exp\Big\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} + \mathbf{L}(\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)^T(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} + \mathbf{L}(\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)$$
$$+ n\lambda(\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)^T\mathbf{\Omega}(\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)\Big\}, \qquad (3.54)$$

due to the proof of Theorem 3.2. Similarly, as in Section 3.5, the likelihood function of $\mathbf{y}$ is

$$p_\beta(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} + \mathbf{L}(\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)^T(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} + \mathbf{L}(\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)\right].$$

We then set a prior for β to be

$$f(\beta) \propto \exp\left\{-\frac{n\lambda}{2\sigma^2}(\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)^T\mathbf{\Omega}(\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)\right\}$$

in order to incorporate the constraint. The joint density for $\mathbf{y}$ and β is then proportional to

$$\begin{aligned}
&\exp\Big\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} + \mathbf{L}(\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)^T(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} + \mathbf{L}(\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta) \\
&+ n\lambda(\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)^T\mathbf{\Omega}(\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\beta)\Big\} \\
=&\exp\Big\{-\frac{1}{2}\left(\beta - \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^{-}\left[\tilde{\mathbf{L}}^T\mathbf{y} - \tilde{\mathbf{L}}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r}\right]\right)^T \\
&\cdot \frac{\left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]}{\sigma^2}\left(\beta - \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^{-}\left[\tilde{\mathbf{L}}^T\mathbf{y} - \tilde{\mathbf{L}}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - (\mathbf{I} - \mathbf{P}_{\mathbf{R}})\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r}\right]\right)\Big\} \\
&\cdot \exp\left\{-\frac{1}{2\sigma^2}g(\mathbf{y})\right\},
\end{aligned}$$

where

$$g(\mathbf{y}) = \left(\mathbf{y}^T\mathbf{y} - 2\mathbf{L}\mathbf{R}^{-}\mathbf{r}\mathbf{y}^T + (\mathbf{L}\mathbf{R}^{-}\mathbf{r})^T(\mathbf{L}\mathbf{R}^{-}\mathbf{r}) + (\mathbf{R}^{-}\mathbf{r})^T\mathbf{\Omega}(\mathbf{R}^{-}\mathbf{r})\right)$$

$$
\begin{aligned}
&- \left[\tilde{\mathbf{L}}^T\mathbf{y} - \tilde{\mathbf{L}}^T\mathbf{L}\mathbf{R}^-\mathbf{r} - (\mathbf{I} - \mathbf{P_R})\boldsymbol{\Omega}\mathbf{R}^-\mathbf{r}\right]^T \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^- \\
&\cdot \left[\tilde{\mathbf{L}}^T\mathbf{y} - \tilde{\mathbf{L}}^T\mathbf{L}\mathbf{R}^-\mathbf{r} - (\mathbf{I} - \mathbf{P_R})\boldsymbol{\Omega}\mathbf{R}^-\mathbf{r}\right].
\end{aligned}
$$

The resulting posterior distribution is a ($p-$ rank($\mathbf{R}$))-variate normal distribution with the posterior mean

$$
\mathrm{E}(\beta|\mathbf{y}) = \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^- \left[\tilde{\mathbf{L}}^T\mathbf{y} - \tilde{\mathbf{L}}^T\mathbf{L}\mathbf{R}^-\mathbf{r} - (\mathbf{I} - \mathbf{P_R})\boldsymbol{\Omega}\mathbf{R}^-\mathbf{r}\right]
$$

and the posterior variance-covariance matrix

$$
\mathrm{Cov}(\beta|\mathbf{y}) = \sigma^2 \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^- .
$$

The forms of credible intervals come from the same methods discussed for unconstrained credible intervals. We define the credible intervals for the constrained problem in the next Definition.

Definition 3.1. For inequality constraints $\mathbf{G}\beta \geq \mathbf{h}$, with active equality constraints $\mathbf{R}\beta = \mathbf{r}$, the $100(1-\alpha)\%$ Bayesian credible interval for $\mu(t_i)$ is

$$
\mu_\lambda(t_i) \pm z_{\alpha/2}\sigma\sqrt{\mathbf{l}_i^T \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^- \mathbf{l}_i} \tag{3.55}
$$

and a $100(1-\alpha)\%$ Bayesian credible interval for $f(\tau)$ is

$$
f_\lambda(\tau) \pm z_{\alpha/2}\sigma\sqrt{\mathbf{x}^T(\tau) \left[\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} + n\lambda\tilde{\boldsymbol{\Omega}}\right]^- \mathbf{x}(\tau)}, \tag{3.56}
$$

where $\mu_\lambda(t_i)$ and $f_\lambda(\tau)$ are constrained estimates of $\mu(t_i)$ and $f(\tau)$, respectively, obtained via formulas (3.48) and (3.52) using the constrained estimate β_λ, $\mathbf{l}_i^T$ is the i-th row vector of $\mathbf{L}$, $\mathbf{x}(\tau)$ is defined as in (3.51), $\mathbf{P_R} = \mathbf{R}^-\mathbf{R}$, $\tilde{\mathbf{L}} = \mathbf{L}(\mathbf{I} - \mathbf{P_R})$, and $\tilde{\boldsymbol{\Omega}} = (\mathbf{I} - \mathbf{P_R})\boldsymbol{\Omega}(\mathbf{I} - \mathbf{P_R})$.

3.6 Proofs

3.6.1 Proof of Theorem 3.1

Proof. We adapt the proof by Lawson and Hanson (1974). First define

$$\begin{aligned}\mathbf{u} &= \mathbf{S}^T\mathbf{b} \\ &= \begin{bmatrix}\mathbf{u}_1 \\ \mathbf{u}_2\end{bmatrix}\begin{matrix}\} & r \\ \} & p-r\end{matrix} \quad ,\end{aligned}$$

and

$$\begin{aligned}\mathbf{v} &= \mathbf{Q}^T\mathbf{Y} \\ &= \begin{bmatrix}\mathbf{v}_1 \\ \mathbf{v}_2\end{bmatrix}\begin{matrix}\} & r \\ \} & n+p-r\end{matrix} \quad .\end{aligned}$$

The invariance property of the Euclidean length for orthogonal matrices gives us

$$\begin{aligned}\| \mathbf{Y} - \mathbf{A}(\lambda)\mathbf{b} \|^2 &= \left\| \mathbf{Y} - \mathbf{Q}\begin{bmatrix}\mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{S}^T\mathbf{b} \right\|^2 \\ &= \left\| \mathbf{Q}^T\mathbf{Y} - \begin{bmatrix}\mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{S}^T\mathbf{b} \right\|^2 \\ &= \left\| \begin{bmatrix}\mathbf{v}_1 \\ \mathbf{v}_2\end{bmatrix} - \begin{bmatrix}\mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 \\ \mathbf{u}_2\end{bmatrix} \right\|^2 \\ &= ||\mathbf{v}_1 - \mathbf{R}_{11}\mathbf{u}_1||^2 + ||\mathbf{v}_2||^2 . \end{aligned} \tag{3.57}$$

The term $\mathbf{v}_2^T\mathbf{v}_2$ in (3.57) is a constant. The unique solution $\mathbf{u}_{\lambda 1}$ of

$$\mathbf{R}_{11}\mathbf{u}_1 = \mathbf{v}_1 \tag{3.58}$$

exists since $\mathbf{R}_{11}$ is full rank. Then all solutions to the problem in (3.57) are of the form

$$\mathbf{u}_\lambda = \begin{bmatrix} \mathbf{u}_{\lambda 1} \\ \mathbf{u}_2 \end{bmatrix}, \tag{3.59}$$

where $\mathbf{u}_2$ is arbitrary. The minimum length vector $\mathbf{u}_\lambda$ corresponds to $\mathbf{u}_2 = \mathbf{0}$. Therefore, the unique solution of minimum length is

$$\mathbf{b}_\lambda = \mathbf{S} \begin{bmatrix} \mathbf{u}_{\lambda 1} \\ \mathbf{0} \end{bmatrix}. \tag{3.60}$$

□

3.6.2 Proof of Theorem 3.2

Proof. Let $\mathbf{R}^-$ represent the generalized inverse of $\mathbf{R}$ and observe that $\mathbf{P_R}$ and $\mathbf{I} - \mathbf{P_R}$ are symmetric and idempotent. Now, for any $\mathbf{b}$ satisfying $\mathbf{Rb} = \mathbf{r}$ we have

$$\begin{aligned} \mathbf{b} &= \mathbf{P_R b} + (\mathbf{I} - \mathbf{P_R})\mathbf{b} \\ &= \mathbf{R}^-\mathbf{Rb} + (\mathbf{I} - \mathbf{P_R})\mathbf{b} \\ &= \mathbf{R}^-\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}. \end{aligned} \tag{3.61}$$

Using this representation the minimization criterion (3.28) becomes

$$\begin{aligned} \mathbf{J}(\mathbf{b}) = & \left[\mathbf{y} - \mathbf{LR}^-\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}\right]^T \left[\mathbf{y} - \mathbf{LR}^-\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}\right] \\ & + n\lambda \left[\mathbf{R}^-\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}\right]^T \boldsymbol{\Omega} \left[\mathbf{R}^-\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}\right]. \end{aligned} \tag{3.62}$$

Now we differentiate $\mathbf{J}(\mathbf{b})$ with respect to $\mathbf{b}$ to obtain

$$\begin{aligned} \frac{1}{2}\frac{\partial \mathbf{J}(\mathbf{b})}{\partial \mathbf{b}} = & -(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{y} + (\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{LR}^-\mathbf{r} \\ & + (\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b} + n\lambda(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}\mathbf{R}^-\mathbf{r} \\ & + n\lambda(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I} - \mathbf{P_R})\mathbf{b}, \end{aligned}$$

which gives the normal equations

$$\left[(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I}-\mathbf{P_R})+n\lambda(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I}-\mathbf{P_R})\right]\mathbf{b}$$
$$=(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\left[\mathbf{L}^{\mathbf{T}}\mathbf{y}-\mathbf{L}^{\mathbf{T}}\mathbf{L}\mathbf{R}^{-}\mathbf{r}-n\lambda\boldsymbol{\Omega}\mathbf{R}^{-}\mathbf{r}\right] \tag{3.63}$$

or, equivalently,

$$\left[(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I}-\mathbf{P_R})+n\lambda(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I}-\mathbf{P_R})\right](\mathbf{I}-\mathbf{P_R})\mathbf{b}$$
$$=(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\left[\mathbf{L}^{\mathbf{T}}\mathbf{y}-\mathbf{L}^{\mathbf{T}}\mathbf{L}\mathbf{R}^{-}\mathbf{r}-n\lambda\boldsymbol{\Omega}\mathbf{R}^{-}\mathbf{r}\right] \tag{3.64}$$

and, hense

$$\mathbf{b}_\lambda=\left[(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I}-\mathbf{P_R})+n\lambda(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I}-\mathbf{P_R})\right]^{-}$$
$$\cdot(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\left[\mathbf{L}^{\mathbf{T}}\mathbf{y}-\mathbf{L}^{\mathbf{T}}\mathbf{L}\mathbf{R}^{-}\mathbf{r}-n\lambda\boldsymbol{\Omega}\mathbf{R}^{-}\mathbf{r}\right]+\mathbf{R}^{-}\mathbf{r}$$

Let $\mathbf{b}_g=\mathbf{b}_\lambda-\mathbf{R}^{-}\mathbf{r}$. We observe the following facts.

(F1) $\mathbf{b}_g$ is invariant with respect to multiplication of $(\mathbf{I}-\mathbf{P_R})$ since

$$(\mathbf{I}-\mathbf{P_R})\left[(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I}-\mathbf{P_R})+n\lambda(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I}-\mathbf{P_R})\right]^{-}$$
$$=\left[\{(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I}-\mathbf{P_R})+n\lambda(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I}-\mathbf{P_R})\}(\mathbf{I}-\mathbf{P_R})\right]^{-}$$
$$=\left[(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I}-\mathbf{P_R})+n\lambda(\mathbf{I}-\mathbf{P_R})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{I}-\mathbf{P_R})\right]^{-},$$
$$(\mathbf{I}-\mathbf{P_R})^{-}=(\mathbf{I}-\mathbf{P_R})$$

and

$$(\mathbf{I}-\mathbf{P_R})^{\mathbf{2}}=(\mathbf{I}-\mathbf{P_R}).$$

(F2) Using (F1), we have

$$\mathbf{R}\mathbf{b}_g=\mathbf{R}(\mathbf{I}-\mathbf{P_R})\mathbf{b}_g$$

$$= (\mathbf{R} - \mathbf{R}\mathbf{R}^{-}\mathbf{R})\mathbf{b}_g$$
$$= \mathbf{0}.$$

(F3) $\mathbf{b}_\lambda = \mathbf{b}_g + \mathbf{R}^{-}\mathbf{r}$ is consistent with $\mathbf{R}\mathbf{b} = \mathbf{r}$. (F3) results from (F2) along with the fact that

$$\begin{aligned}\mathbf{R}\mathbf{b}_\lambda &= \mathbf{0} + \mathbf{R}\mathbf{R}^{-}\mathbf{r}\\ &= \mathbf{r},\end{aligned}$$

and

$$\left[(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I} - \mathbf{P_R}) + n\lambda(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{\Omega}(\mathbf{I} - \mathbf{P_R})\right]\mathbf{R}^{-}\mathbf{r} = 0$$

since

$$(\mathbf{I} - \mathbf{P_R})\mathbf{R}^{-} = \mathbf{0}.$$

As a result of (F1)-(F3),

$$\begin{aligned}\mathbf{b}_\lambda &= \left[(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{L}^{\mathbf{T}}\mathbf{L}(\mathbf{I} - \mathbf{P_R}) + n\lambda(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\mathbf{\Omega}(\mathbf{I} - \mathbf{P_R})\right]^{-}\{(\mathbf{I} - \mathbf{P_R})^{\mathbf{T}}\\ &\quad \cdot (\mathbf{L}^{\mathbf{T}}\mathbf{y} - \mathbf{L}^{\mathbf{T}}\mathbf{L}\mathbf{R}^{-}\mathbf{r} - n\lambda\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r})\} + \mathbf{R}^{-}\mathbf{r}\\ &= \left[\tilde{\mathbf{L}}^{\mathbf{T}}\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^{-}\tilde{\mathbf{L}}^{\mathbf{T}}(\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r}) - n\lambda\left[\tilde{\mathbf{L}}^{\mathbf{T}}\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^{-}(\mathbf{I} - \mathbf{P_R})\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r} + \mathbf{R}^{-}\mathbf{r}\\ &= \left[\tilde{\mathbf{L}}^{T}\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^{-}\tilde{\mathbf{L}}^{T}\mathbf{y} - \left[\tilde{\mathbf{L}}^{T}\tilde{\mathbf{L}} + n\lambda\tilde{\mathbf{\Omega}}\right]^{-}(\mathbf{I} - \mathbf{P_R})\left[\mathbf{L}^{T}\mathbf{L} + n\lambda\mathbf{\Omega}\right]\mathbf{R}^{-}\mathbf{r} + \mathbf{R}^{-}\mathbf{r}.\end{aligned}$$

To prove that $\mathbf{b}_\lambda$ achieves the global minimum of $\mathbf{J}(\mathbf{b})$ in (3.62) it is enough to show that $\mathbf{J}(\hat{\mathbf{b}}) \geq \mathbf{J}(\mathbf{b}_\lambda)$, where $\hat{\mathbf{b}}$ is any estimator of $\mathbf{b}$ with $\mathbf{R}\hat{\mathbf{b}} = \mathbf{r}$. For this purpose, observe that

$$\begin{aligned}\mathbf{J}(\hat{\mathbf{b}}) &= \left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\hat{\mathbf{b}}\right]^{T}\left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\hat{\mathbf{b}}\right]\\ &\quad + n\lambda\left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\hat{\mathbf{b}}\right]^{T}\mathbf{\Omega}\left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\hat{\mathbf{b}}\right]\\ &= \left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda + \mathbf{L}(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]^{T}\end{aligned}$$

$$
\begin{aligned}
&\cdot \left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda + \mathbf{L}(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\
&+ n\lambda \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda - (\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]^T \mathbf{\Omega} \\
&\cdot \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda - (\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\
= &\left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right] && (3.65) \\
&+ 2\left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{L}(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}}) && (3.66) \\
&+ \left[\mathbf{L}(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]^T \left[\mathbf{L}(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] && (3.67) \\
&+ n\lambda \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{\Omega} \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right] && (3.68) \\
&+ n\lambda \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]^T \mathbf{\Omega} \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] && (3.69) \\
&- 2n\lambda \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{\Omega} \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]. && (3.70)
\end{aligned}
$$

Now observe that summation of the two cross-product terms in (3.66) and (3.70) is 0. To see this write

$$
\begin{aligned}
\mathbf{S}(\mathbf{b}_\lambda, \hat{\mathbf{b}}) &= \left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{L}(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}}) \\
&\quad - n\lambda \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{\Omega} \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\
&= \left[\left[\mathbf{y} - \mathbf{L}\mathbf{R}^{-}\mathbf{r} - \mathbf{L}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{L} - n\lambda \left[\mathbf{R}^{-}\mathbf{r} + (\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \mathbf{\Omega}\right] \\
&\quad \cdot \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\
&= \left[\left[\mathbf{L}^T\mathbf{y} - \mathbf{L}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - n\lambda\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r}\right]^T - \left[\{\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T\right] \\
&\quad \cdot \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\
&= \Big[\left[(\mathbf{I} - \mathbf{P_R})\{\mathbf{L}^T\mathbf{y} - \mathbf{L}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - n\lambda\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r}\}\right]^T \\
&\quad - \left[(\mathbf{I} - \mathbf{P_R})\{\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda\right]^T \Big] \left[(\mathbf{I} - \mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\
&= 0,
\end{aligned}
$$

since $\mathbf{b}_\lambda$ is the solution of the normal equation in (3.64) such that

$$(\mathbf{I} - \mathbf{P_R})\{\mathbf{L}^T\mathbf{L} + n\lambda\mathbf{\Omega}\}(\mathbf{I} - \mathbf{P_R})\mathbf{b}_\lambda = (\mathbf{I} - \mathbf{P_R})\left[\mathbf{L}^T\mathbf{y} - \mathbf{L}^T\mathbf{L}\mathbf{R}^{-}\mathbf{r} - n\lambda\mathbf{\Omega}\mathbf{R}^{-}\mathbf{r}\right].$$

Finally, we have

$$\begin{aligned}\mathbf{J}(\hat{\mathbf{b}}) &= \mathbf{J}(\mathbf{b}_\lambda) + \left[\mathbf{L}(\mathbf{I}-\mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]^T \left[\mathbf{L}(\mathbf{I}-\mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\ &\quad + n\lambda \left[(\mathbf{I}-\mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right]^T \boldsymbol{\Omega} \left[(\mathbf{I}-\mathbf{P_R})(\mathbf{b}_\lambda - \hat{\mathbf{b}})\right] \\ &\geq \mathbf{J}(\mathbf{b}_\lambda),\end{aligned}$$

and equality holds only if $\mathbf{b}_\lambda = \hat{\mathbf{b}}$. Therefore, $\mathbf{b}_\lambda$ is unique and achieves the global minimum. $\square$

CHAPTER IV

FINITE SAMPLE STUDIES

In this chapter we investigate various issues in ill-posed inverse problems discussed in the previous chapters based on simulation studies. We explore several ill-posed inverse problems using different combinations of kernel functions, inverse functions and levels of errors. In Section 4.1, we conduct empirical investigations to give the finite sample behavior of an MOR estimator. The remaining sections include empirical investigations for the results in Chapter III. We examine the performance of $GCV_C(\lambda)$ with empirical investigations of the Bayesian credible intervals in Section 4.2. In Section 4.3, we show some results obtained by applying our work to a NMR experimental data analysis setting.

The accuracy of estimates of the mean function and the inverse function highly depend on a good choice of the smoothing parameter along with very stable and accurate computations. Using the properties of types of banded matrices based on the B-spline basis functions, the speed of computation is improved eliminating unnecessary computations. For the purpose of efficient numerical integration, *Gauss-Kronrod* formulas are adopted (Press, Teukolsky, Vetterling, and Flannery, 1986). Numerical integration based on Gauss-Kronrod formulas is an adaptive quadrature that finds some number of Gaussian nodes and adds some more nodes chosen to obtain the highest possible degree of precision. We conduct various simulation studies on equally spaced design points and knots throughout this Chapter.

4.1 Simulation Design

Our simulation studies are conducted on combinations of various types of kernel functions and inverse functions with different levels of errors. We control the frequencies of the true inverse functions using forms of sine functions. We define two types of the true inverse functions and two types of kernel functions as follows:

$$
\begin{aligned}
f^L(\tau) &= \frac{\sin 2\pi(\tau - \frac{1}{2})/2\pi(\tau - \frac{1}{2})}{\int_0^1 \sin 2\pi(\tau - \frac{1}{2})/2\pi(\tau - \frac{1}{2})\ d\tau}, \\
f^H(\tau) &= \frac{0.22 + \sin 11\pi(\tau - \frac{1}{2})/11\pi(\tau - \frac{1}{2})}{0.22 + \int_0^1 \sin 11\pi(\tau - \frac{1}{2})/11\pi(\tau - \frac{1}{2})\ d\tau}, \\
K^L(t,\tau) &= \frac{1}{2}\left[f_k^*(t-\tau, 0.15) + f_k^*(t+\tau, 0.15)\right]
\end{aligned}
$$

and

$$
K^H(t,\tau) = \frac{1}{2}\left[f_k^*(t-\tau, 0.05) + f_k^*(t+\tau, 0.05)\right],
$$

where

$$
f_k^*(u,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{u-m}{\sigma}\right)^2\right\},
$$

$$
m = \begin{cases} \frac{3}{2} & \text{if } 1 < u \leq 2, \\ \frac{1}{2} & \text{if } 0 < u \leq 1, \\ -\frac{1}{2} & \text{if } -1 \leq u \leq 0, \end{cases}
$$

for $\tau, t \in [0,1]$. The upper case letters L and H stand for low frequency and high frequency, respectively.

Figure 2 shows the true inverse functions and $f_k^*(u, 0.15)$ and $f_k^*(u, 0.05)$ for $u \in [0,1]$. The kernel functions $K^L(t,\tau)$ and $K^H(t,\tau)$ are presented in figure 3. We see how frequencies of the products of the true inverse functions and kernel functions are affected by the frequencies of the inverse functions and kernel functions in figure 4.

The diagrams in Figure 5 show the true mean functions that correspond to the cases in figure 4. Figure 5 (a) shows an example of a mean function that reflects the low frequencies of the inverse function f and kernel function K. When the K^L kernel function is employed with f^H as in figure 5 (b), the kernel function represents a function that is too smooth relative to the inverse function. In this case, the kernel function can smear out the frequencies of f and the mean function becomes low frequency in nature. This is a case when effective deconvolution can be quite difficult.

On the other hand, figure 5 (c) shows the case when a mean function has relatively high frequencies due to the high frequencies of the kernel function even though the inverse function has low frequencies. This is a case when the frequencies of the mean function are quite different from the frequencies of the inverse functions. In this case the optimal smoothing parameters λ_f and λ_μ can be quite different. We will define λ_f and λ_μ in the following section in our simulation settings.

First we investigate the patterns of Fourier coefficients for the inverse functions and the kernel functions. Let f_j^L and f_j^H be the Fourier coefficients of $f^L(\tau)$ and $f^H(\tau)$ for $j = 1, 2, \ldots$ Similarly k_j^L and k_j^H are defined for $K^L(t, \tau)$ and $K^H(t, \tau)$. Figure 6 shows the obtained absolute values of the Fourier coefficients. We observe that $|f_j^*|$ and $|k_j^*| \to 0$ as $j \to \infty$, where $* = L$ or H. The rate of decay of the Fourier coefficients depends on the smoothness of each function. Smooth functions have faster decay rate of the Fourier coefficients than functions that are less smooth. Figure 7 shows the Fourier coefficients of the mean function $\mu_j^{**} = k_j^* f_j^*$, where $* = L$ or H.

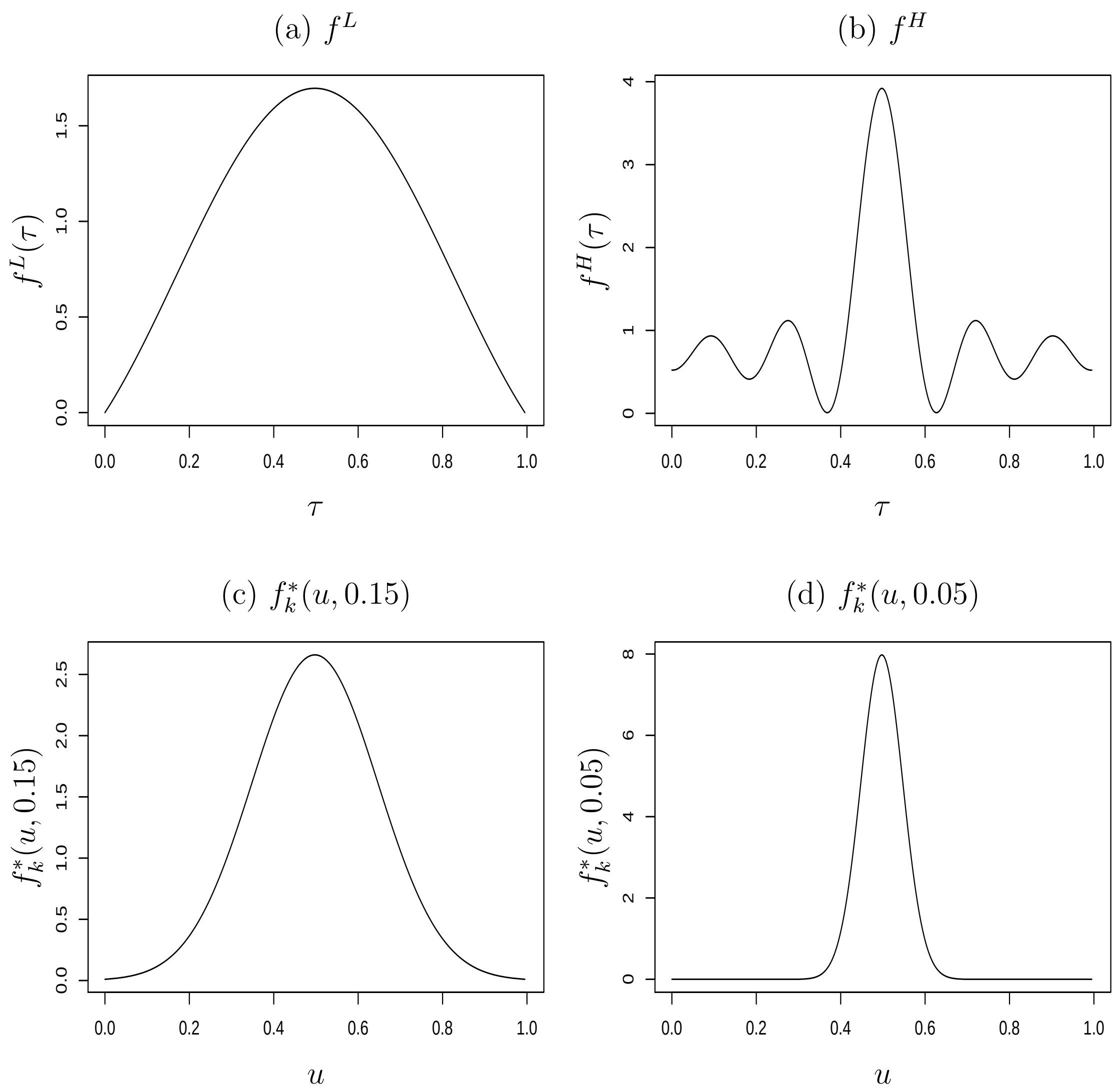

FIGURE 2: True inverse functions and $f_k^*(u, \sigma)$.

(a) K^L (b) K^H

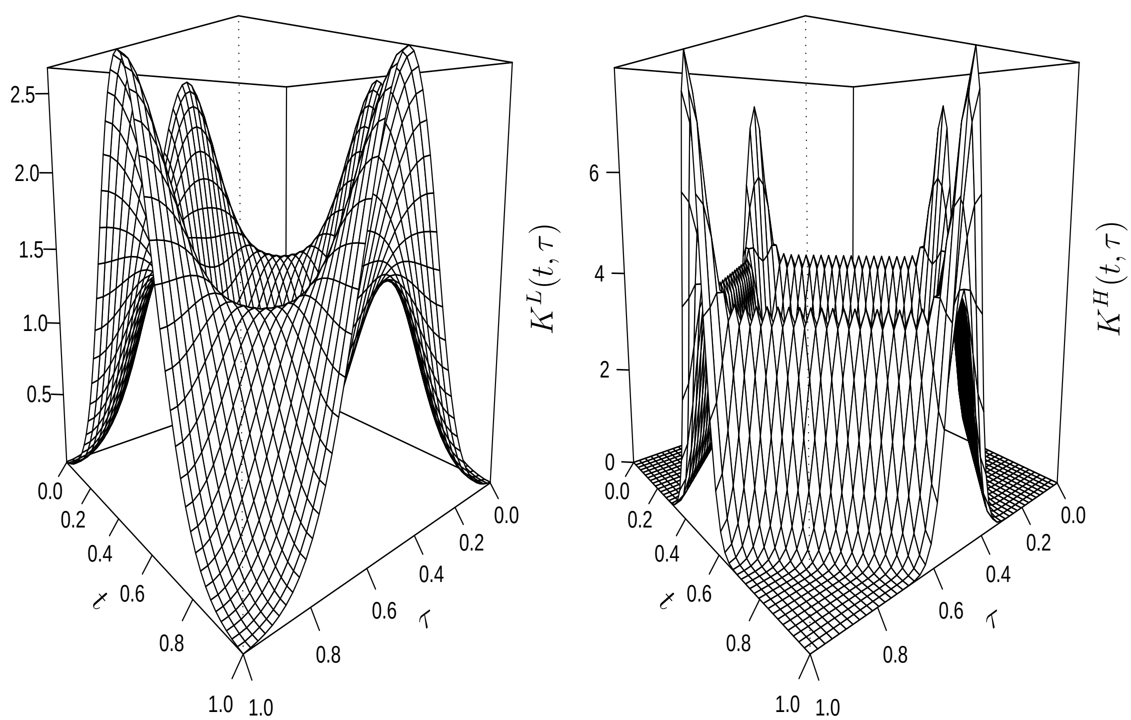

FIGURE 3: Kernel functions.

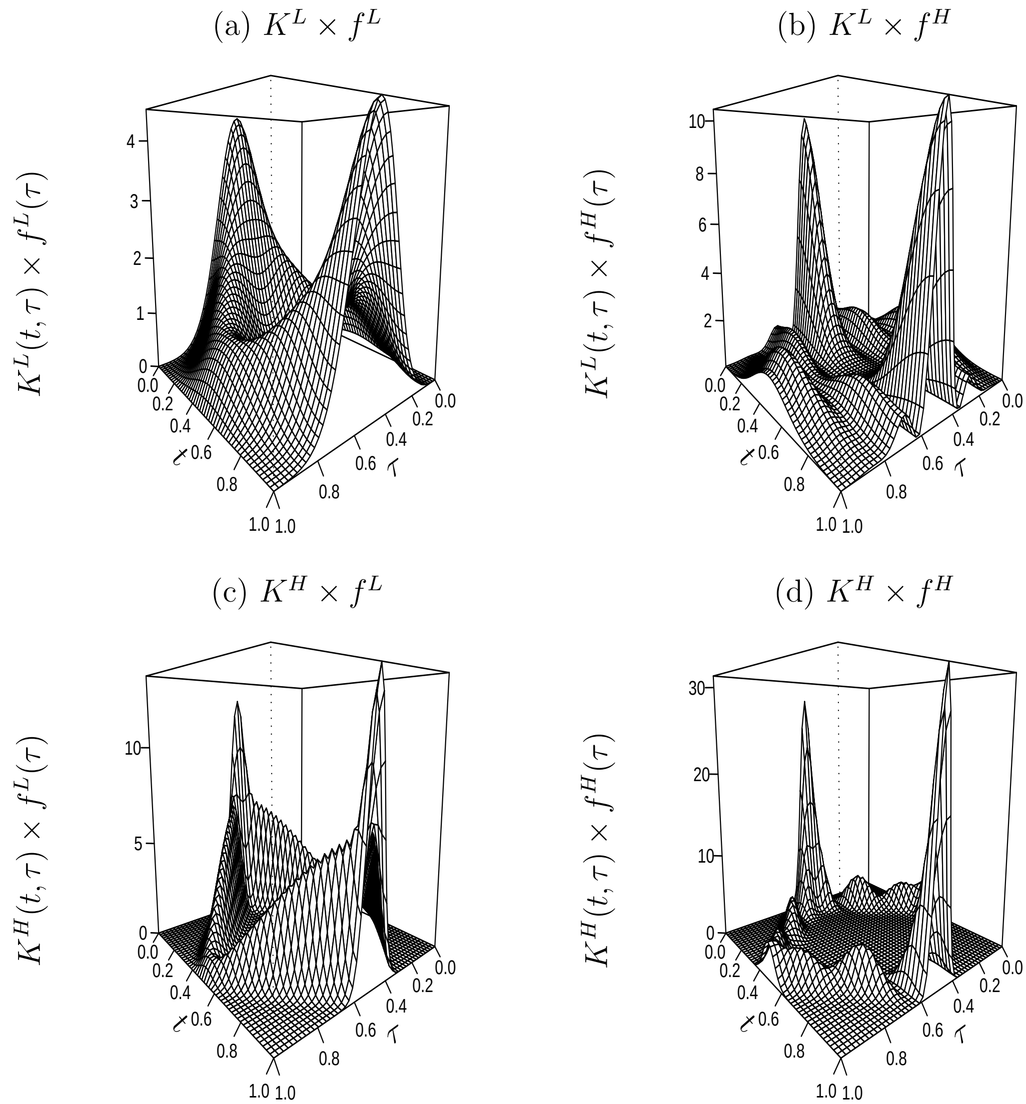

FIGURE 4: Product of true inverse functions and kernel functions.

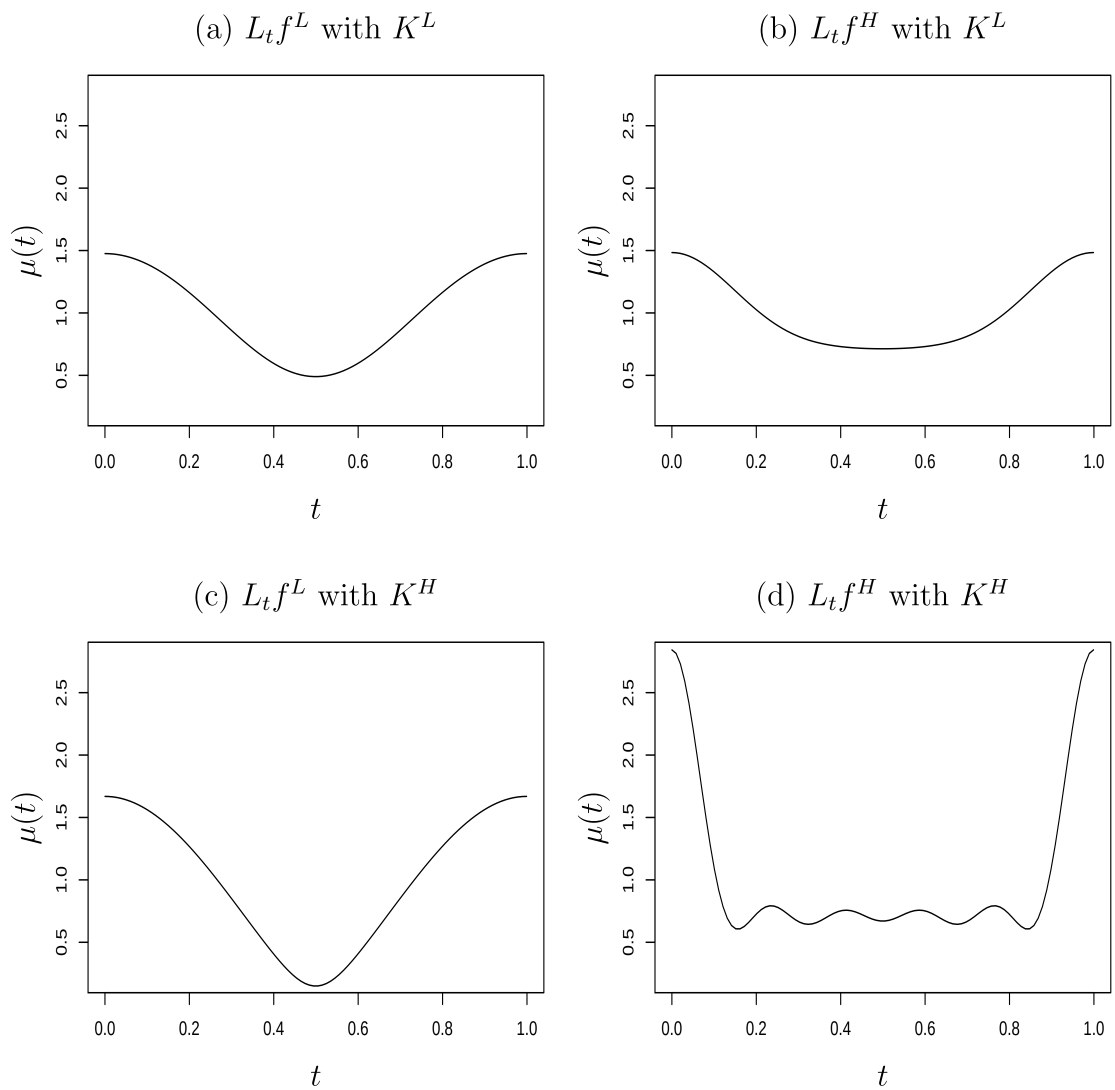

FIGURE 5: True mean functions.

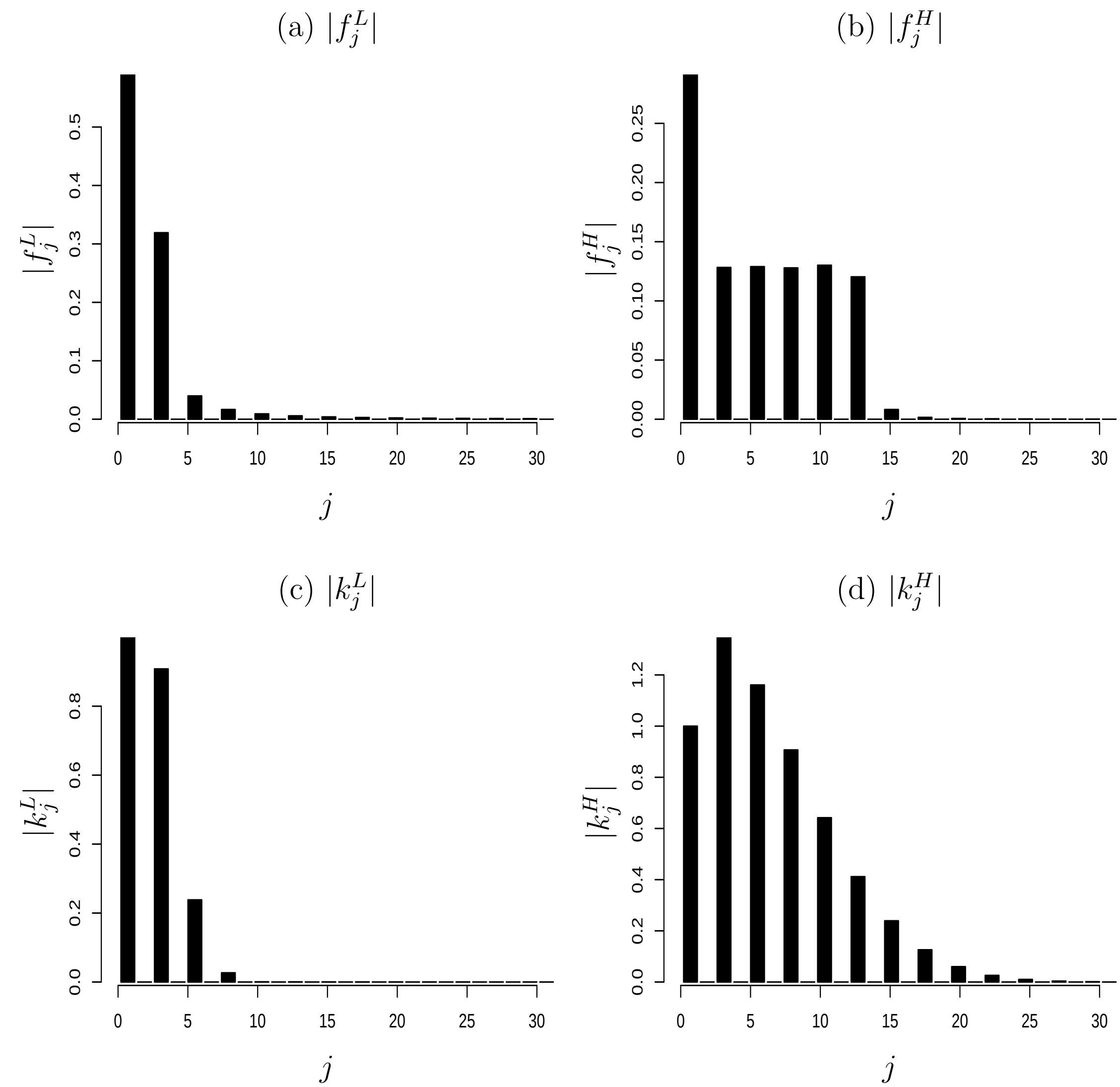

FIGURE 6: Fourier coefficients of inverse functions and kernel functions.

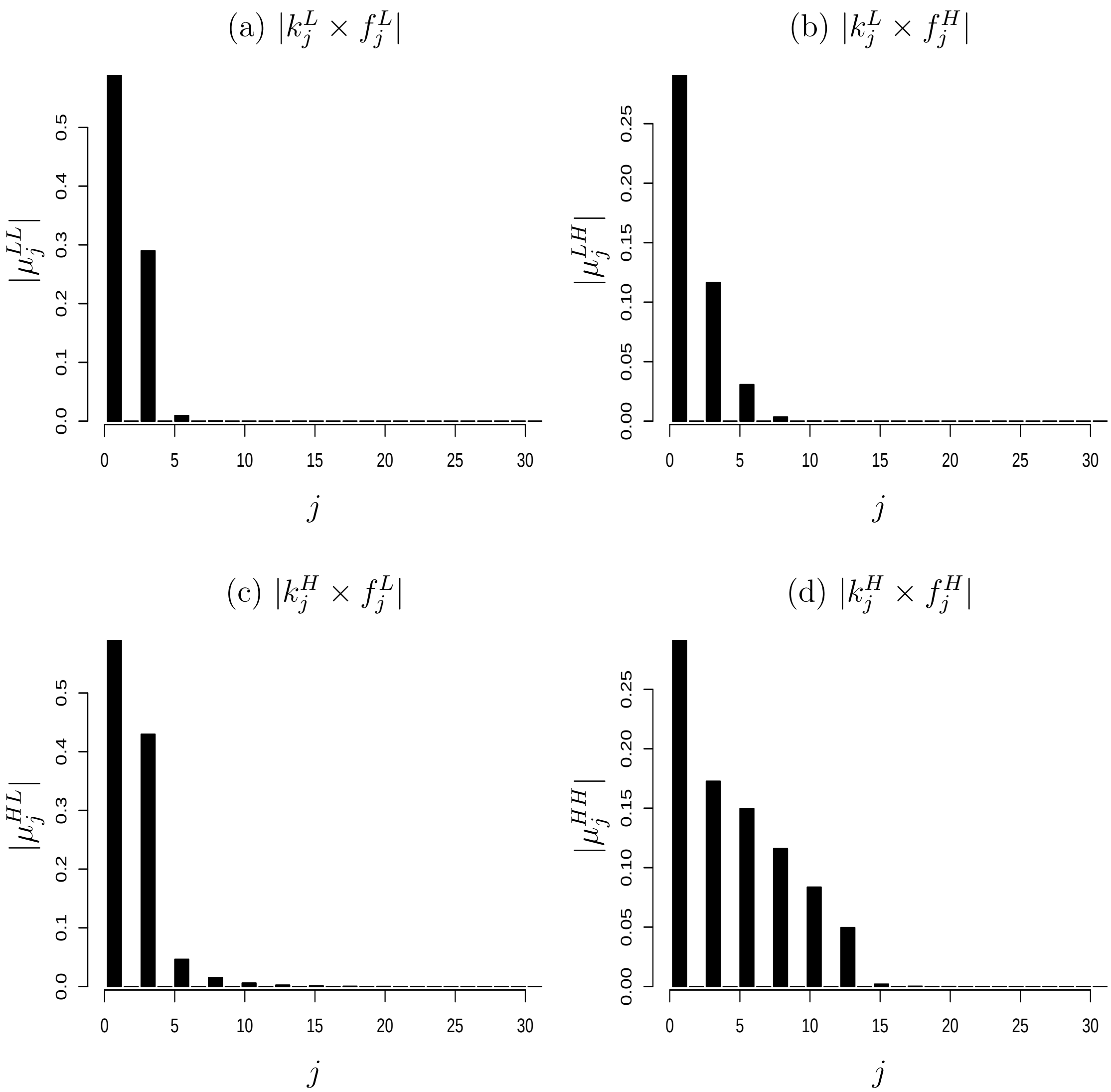

FIGURE 7: Fourier coefficients of mean functions.

4.2 Simulation Results

We generate simulated data sets adding independent normal random errors to the true mean functions on 200 equally spaced design points on $(0,1)$. Let L_μ and L_f be the loss functions for μ and f, respectively. For a given λ, we evaluate the loss functions on the design points, where

$$\mathrm{L}_\mu = \frac{1}{200}\sum_{i=1}^{200}\left[\mu\left(\frac{2i-1}{400}\right) - \mu_\lambda\left(\frac{2i-1}{400}\right)\right]^2 \tag{4.1}$$

and

$$\mathrm{L}_f = \frac{1}{200}\sum_{i=1}^{200}\left[f\left(\frac{2i-1}{400}\right) - f_\lambda\left(\frac{2i-1}{400}\right)\right]^2. \tag{4.2}$$

We find optimal $\tilde{\lambda}_\mu$ and $\tilde{\lambda}_f$ that minimize L_μ and L_f, respectively.

Large sample properties of the behavior of the optimal $\tilde{\lambda}_\mu$ and $\tilde{\lambda}_f$ are examined by increasing the sample size n. More specifically, we generated samples of size $n = 100, 300, 500$ and 1000 and the ratios $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ were calculated over 100 replicate samples. Table 1 presents results, averaged over the 100 replications, for the sample mean and standard deviation of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ in our simulation settings.

In general, the sample mean values of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ are close to constants over sample sizes for each case. However the values of standard deviation of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ are decreasing as n increases showing $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ tends to a constant as the sample size increases. Table 1 gives the information about the behavior of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ depending on the smoothness of K and f. For example, when $K = K^L$ and $f = f^H$ the standard deviation of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ tends to 0 much faster than the case of $K = K^L$ and $f = f^L$. We also discover that the standard deviation of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ for $K = K^H$ and $f = f^H$ converges to 0 much faster than the example of $K = K^H$ and $f = f^L$. These support the results from Theorem 2.2 and Theorem 2.3 in Chapter II by showing that when we have high frequencies of f and a smooth kernel function K, $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ tends to a constant as sample size increases. However,

TABLE 1: Empirical properties of the ratio of $\tilde{\lambda}_\mu$ and $\tilde{\lambda}_f$.

Sample Mean of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$						
σ of noise	K	f	$n = 100$	$n = 300$	$n = 500$	$n = 1000$
.01	K^L	f^L	.0370224	.0372017	.0374177	.0373232
	K^L	f^H	.0022971	.0022880	.0022873	.0022897
	K^H	f^L	.4557705	.4631467	.4646367	.4687726
	K^H	f^H	.6035533	.6403305	.6592470	.6652390
.10	K^L	f^L	.0639122	.0486692	.0427545	.0379822
	K^L	f^H	.0001124	.0001117	.0001116	.0001120
	K^H	f^L	.3290972	.3169472	.2946034	.3416490
	K^H	f^H	.0019995	.0020613	.0020145	.0020114
Standard Deviation of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$						
σ of noise	K	f	$n = 100$	$n = 300$	$n = 500$	$n = 1000$
.01	K^L	f^L	.0037758	.0022264	.0016597	.0012696
	K^L	f^H	.0001454	.0000803	.0000570	.0000464
	K^H	f^L	.0942721	.0551209	.0447745	.0312152
	K^H	f^H	.1603572	.1028730	.0782265	.0548021
.10	K^L	f^L	.0735999	.0447150	.0287205	.0123847
	K^L	f^H	.0000092	.0000015	.0000012	.0000009
	K^H	f^L	.4220915	.2943494	.2468576	.2531637
	K^H	f^H	.0005189	.0002576	.0002097	.0001352

the standard deviation value of $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ is far from 0 when we have high frequencies of K relative to the smoothness of f.

Look at the Fourier series in figure 6 and figure 7 in order to understand the nature of the problem. When we have high frequencies of f and low frequencies of K we observe that the $|f_j|$ decay relatively slower than the $|k_j|$ as j increases. The Fourier coefficients of the mean function satisfy the equation $|\mu_j| = |k_j f_j|$. The Fourier coefficients of mean function $|\mu_j|$ decays relatively slowly when we have high frequencies of f as you can see in figure 7. In this case the frequencies of the mean function are mainly affected by the frequencies of f and $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ is close to a constant.

We now investigate the performance of $\mathrm{GCV}_C(\lambda)$ in (3.44) and the Bayesian credible intervals. Figure 8 shows L_μ in (4.1), L_f in (4.2) and $\mathrm{GCV}_C(\lambda)$, where the true inverse function is f^H, the kernel function is K^H and the noise that are added are from $N(0,(0.1)^2)$. We see that $\mathrm{GCV}_C(\lambda)$ is a good estimator of L_μ since $\mathrm{GCV}_C(\lambda)$ is very close to L_μ for all λ's. Let $\hat{\lambda}$ be the minimizer of $\mathrm{GCV}_C(\lambda)$. In this example, we have the optimal $\tilde{\lambda}_\mu = 4.3821 \times 10^{-9}$, the optimal $\tilde{\lambda}_f = 5.4789 \times 10^{-9}$ and $\mathrm{GCV}_C(\lambda)$ is minimized at $\hat{\lambda} = 1.3826 \times 10^{-9}$. The $\hat{\lambda}$ minimizing $\mathrm{GCV}_C(\lambda)$ is also very close to the optimal $\tilde{\lambda}_f$. This example shows a case that we have $\mathrm{GCV}_C(\lambda)$ as a good estimator of L_μ and L_f.

Using the optimal λ minimizing $\mathrm{GCV}_C(\lambda)$, figure 9 shows the $\mu_{\hat{\lambda}}$ which is the fit to the measurement data corresponding to the mean function and the Bayesian credible interval estimates for $\mu(t)$. We present $f_{\hat{\lambda}}$ the estimate of the inverse function and its Bayesian credible interval estimates in figure 10. The $f_{\hat{\lambda}}(\tau)$ is so close to the true inverse function $f(\tau)$ and the two lines showing fitted values of $f_{\hat{\lambda}}(\tau)$ and $f(\tau)$ may look like one line in the figure except when τ's correspond to peaks or the boundary region.

For practical calculations we explain how we find the optimal λ that minimizes

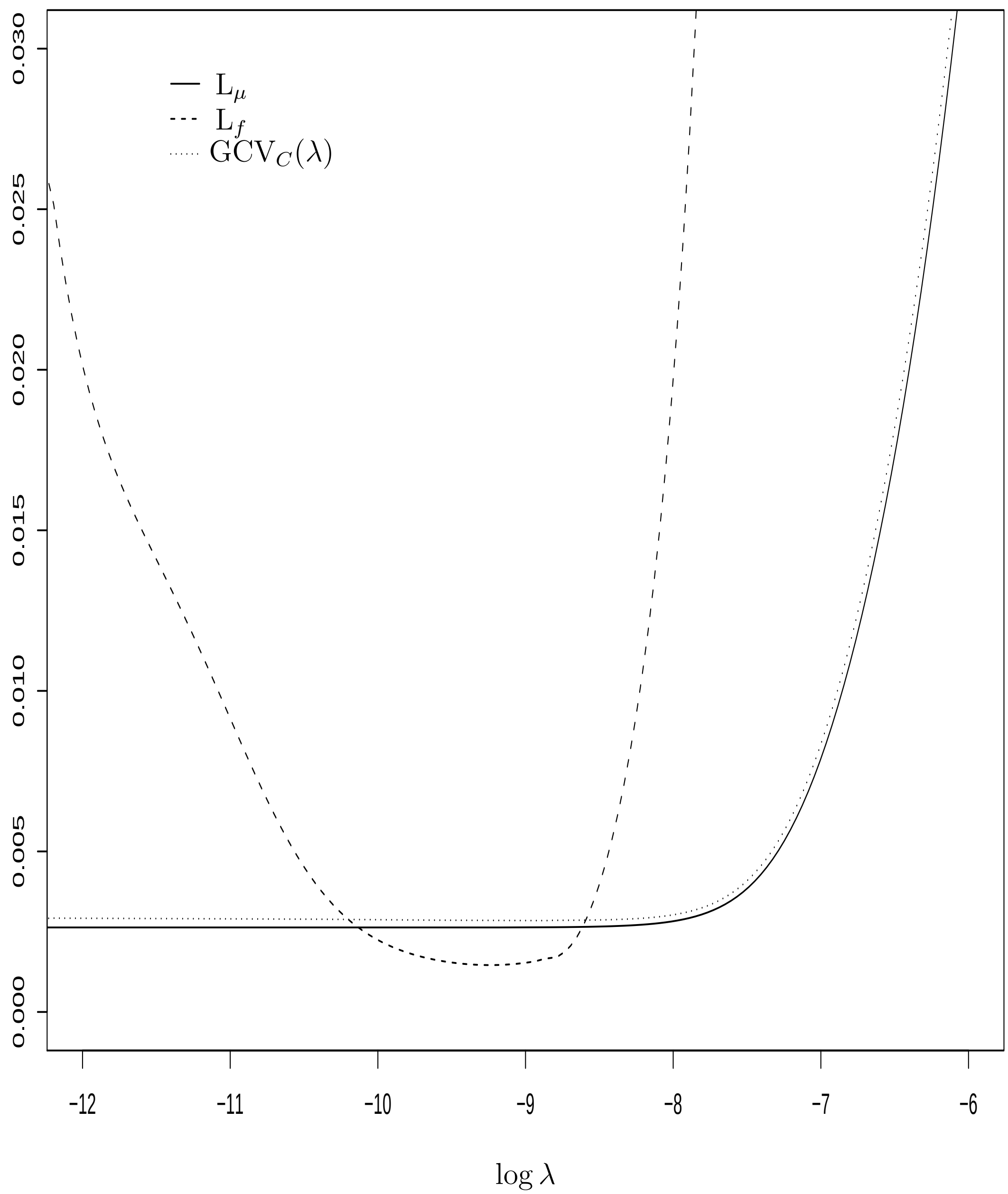

FIGURE 8: L_μ, L_f and $GCV_C(\lambda)$ for a simulated data set for f^H and K^H.

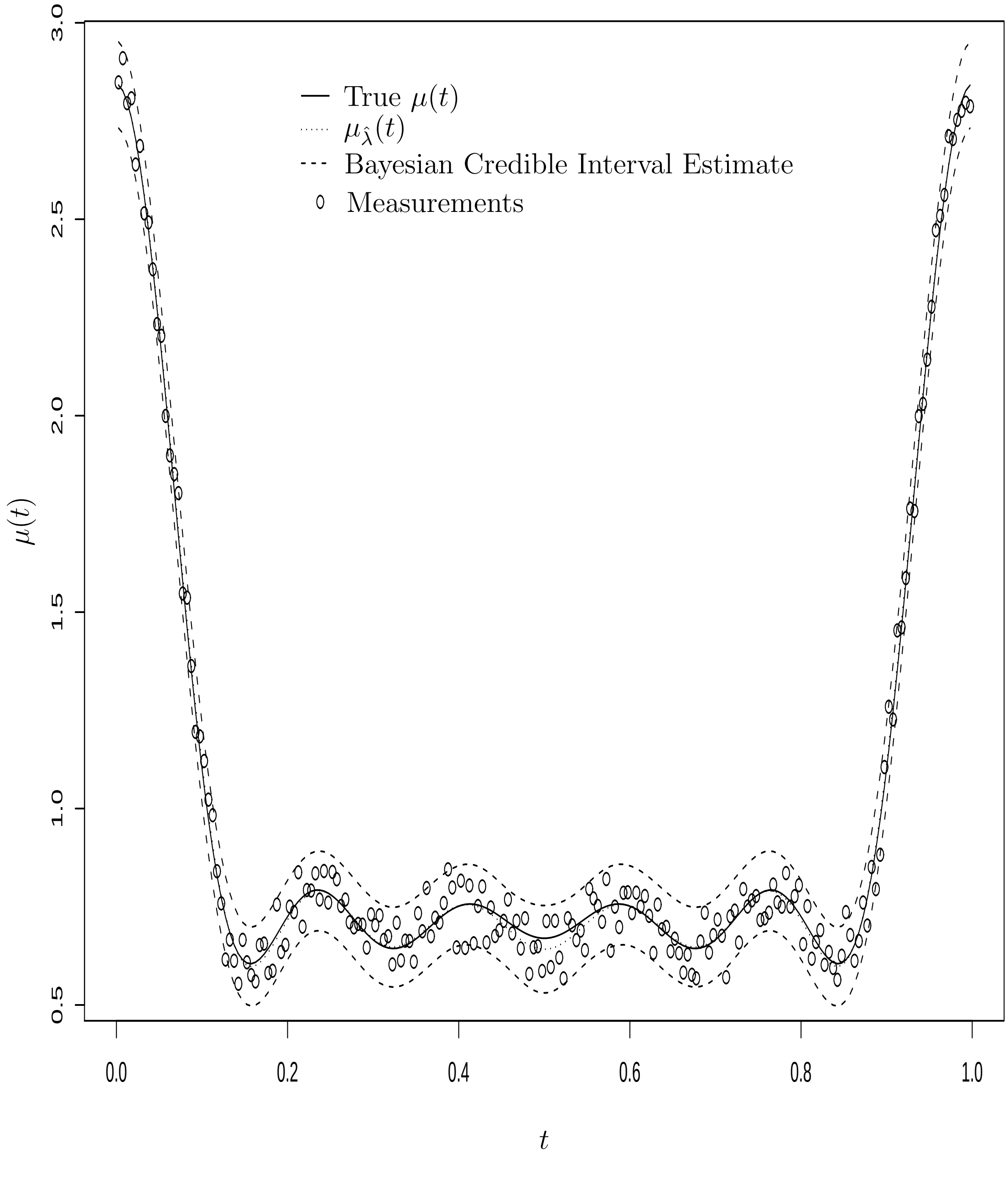

FIGURE 9: $\mu_{\hat{\lambda}}$ and Bayesian credible interval estimates for a simulated data set for f^H and K^H.

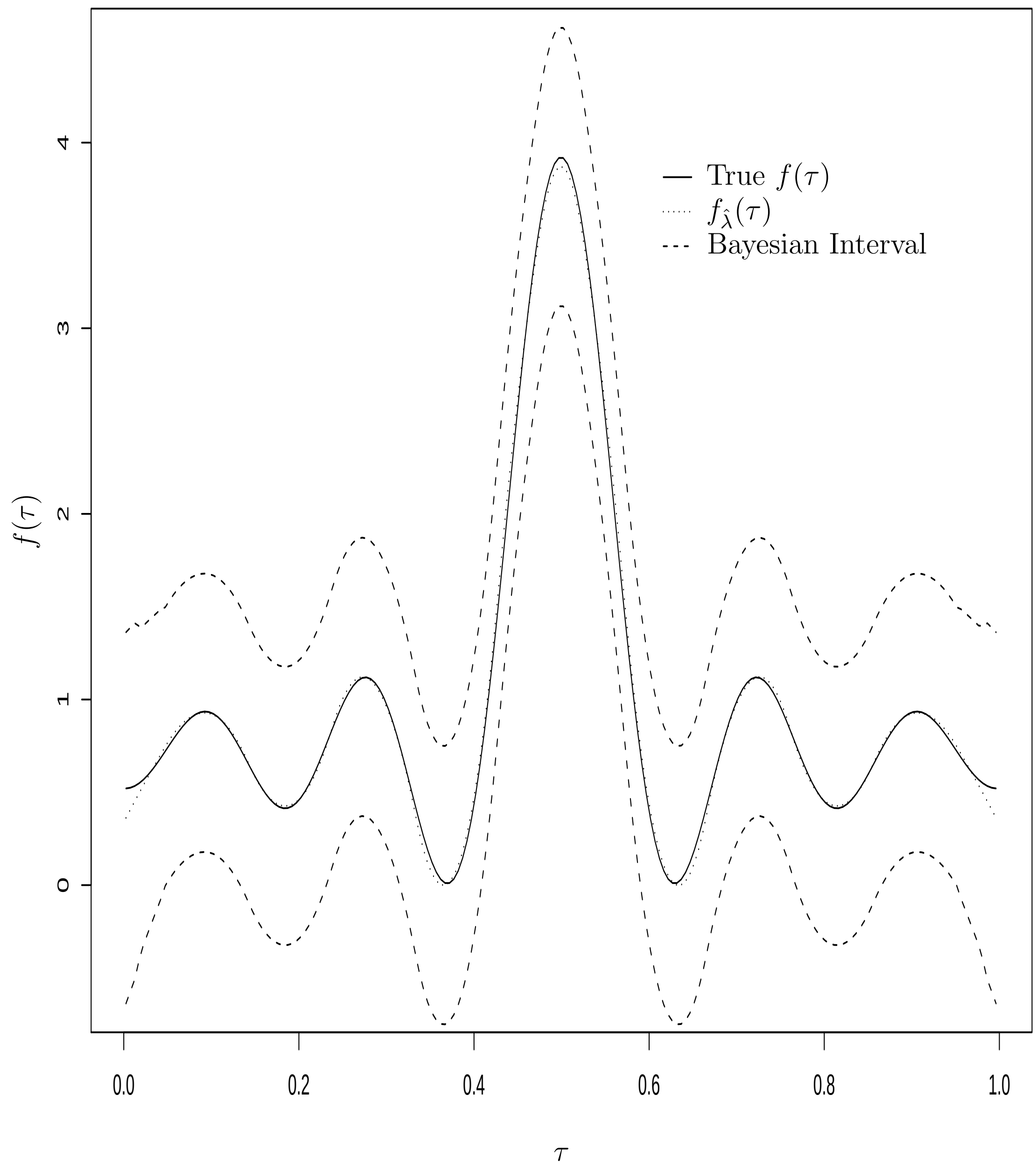

FIGURE 10: $f_{\hat{\lambda}}$ and Bayesian credible interval estimates for a simulated data set for f^H and K^H.

$\mathrm{GCV}_C(\lambda)$. First we make 200 equally spaced partitions on $[e^{-100}, 100]$ based on a log scale so that the ith partition corresponds to $\exp\{-100 + \frac{\log 100 - \log e^{-100}}{200} \times i\}$, $i = 1, \ldots, 200$. In most cases, the minimum bound, e^{-100}, and the maximum bound, 100, for λ are sufficient for our examples. For each of the 200 λ's, we calculate $\mathrm{GCV}_C(\lambda)$. Then we choose three λ's that attain the smallest $\mathrm{GCV}_C(\lambda)$. Then we define some neighborhoods for the three λ's. We employ an automatic search on the neighborhoods of the three λ's. The adopted automatic minimum search method is a combination of a golden section search and successive parabolic interpolations. Then, we have $\hat{\lambda}$ as the minimizer of the $\mathrm{GCV}_C(\lambda)$. The $\hat{\lambda}$ is an estimator of the optimal $\tilde{\lambda}_\mu$ minimizing the loss function in (4.1) for μ or the optimal $\tilde{\lambda}_f$ minimizing the loss function in (4.2) for f. Using the $\hat{\lambda}$ we calculate the estimates of the mean function, the inverse function and the Bayesian credible intervals.

We investigate the empirical properties of Bayesian credible intervals from a frequentist viewpoint, when the smoothing parameters are selected by $GCV_C(\lambda)$. We consider several cases combining K^*, f^*, where $*{=}H$ or L, and levels of errors. We generate 1,000 simulated data sets for each case, independently. For each simulated data set, we select $\hat{\lambda}$ minimizing $GCV_C(\lambda)$. Then we calculate 95% Bayesian credible intervals for μ and f in each case. Our simulated data sets have the same fixed 200 equally spaced design points on $(0, 1)$ for μ and f. We used different values of the seed for the Gaussian random errors in each of the 1,000 replicates. For each replication and every design points, we calculate the overall averages of the covered proportions of the true mean function and the true inverse function by the Bayesian credible intervals.

The average coverage for the Bayesian credible intervals are listed in table 2. In general, we have approximately between 91.66% and 93.97% coverages for μ using 95% Bayesian credible intervals for μ in our simulation settings.

TABLE 2: Average (across the design) coverage for the Bayesian credible intervals for the mean function and the inverse function with 1,000 replications.

σ of noise	K	f	Coverage for μ	Coverage for f
.05	K^L	f^L	.9263	.9844
	K^L	f^H	.9311	.9309
	K^H	f^L	.9166	.9995
	K^H	f^H	.9375	.9965
.01	K^L	f^L	.9257	.9335
	K^L	f^H	.9466	.8537
	K^H	f^L	.9312	.9996
	K^H	f^H	.9397	.9978

Regarding the coverage for f, the coverage ranges from 85.37% to 99.95%. One interesting case is when we have f^H and K^L showing lower coverage than other cases even though it is shown that $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$ behaves like a constant for large sample size as shown in table 1. As a matter of fact, Theorem 2.2 and Theorem 2.3 is not related to the accuracy of estimates or interval estimates rather to the behaviors of the ratio $\frac{\tilde{\lambda}_\mu}{\tilde{\lambda}_f}$. This is a good example indicating that good performance of the Bayesian credible intervals for f depends on the nature of the ill-posed inverse problem. In this case, $|k_j^L|$ decays much faster than the decay rate of $|f_j^H|$. Thus, the low frequencies of K^L damp out the frequencies of f^H and complicate estimating the inverse function. Figure 11 shows the difficulties in estimating the inverse function f^H for a typical data set from the simulation.

Based on our simulation studies we conclude that in order to recover the inverse function f accurately the frequencies of f should not be damped out by the low frequencies of the Kernel function. We also expect a good $\hat{\lambda}$ from $GCV_C(\lambda)$ when we have high frequencies of f. The ideal situation is when both $|k_j f_j|$ and $|f_j|$ decay slowly and similarly.

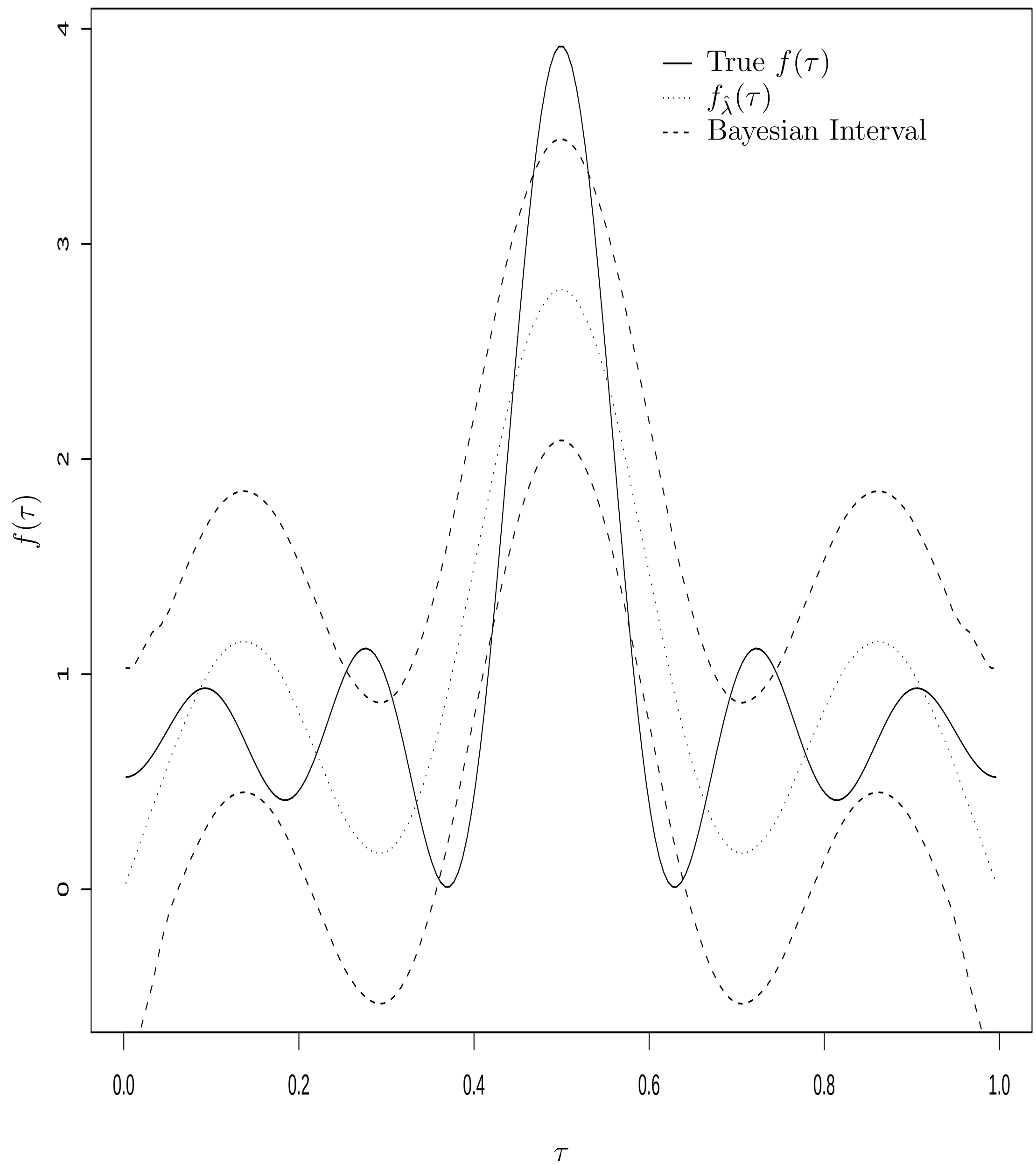

FIGURE 11: $f_{\hat{\lambda}}$ and Bayesian credible interval estimates for a simulated data set for f^H and K^L.

TABLE 3: The optimal λ values minimizing $\mathrm{GCV}_C(\lambda)$ and $\hat{P}_C(\lambda)$ for sample 1 and sample 2.

Experimental Data	$\mathrm{GCV}_C(\lambda)$	$\hat{P}_C(\lambda)$
Sample 1	$e^{-60.47193}$	$e^{-59.68177}$
Sample 2	$e^{-67.57382}$	$e^{-68.08852}$
Sample 3	$e^{-53.63985}$	$e^{-51.27502}$
Sample 4	$e^{-52.88463}$	$e^{-53.25308}$

4.3 NMR Experimental Data Analysis

In this section we apply our estimation methodology to real NMR experimental data from the Engineering Imaging Laboratory at Texas A&M University. For this purpose a set of laboratory experimental data was obtained from the T_1 inversion-recovery measurement conducted on samples of Opalinus clay from Benken, Switzerland. The T_1 measurements were made at 85.6 MHz using the inversion recovery pulse sequence.

The estimates of μ and f using $\mathrm{GCV}_C(\lambda)$ criterion can be found in figure 12 for sample 1 and sample 2.

The $\mathrm{GCV}_C(\lambda)$ and $\hat{P}_C(\lambda)$ criteria give us almost identical estimates of the optimal λ. Table 3 shows the actual values of $\hat{\lambda}$ chosen by these criteria. We have very small values of $\hat{\lambda}$ since our experimental data sets have very small noises. Therefore, the estimates of μ are very similar to lines that are connected with each data point.

The estimates of the relaxation distribution f from sample 1 and sample 2 of Opalinus clay have three modes representing three spectra. Table 4 identifies the locations of τ for the modes of the spectra. Even though we do not have those modes of spectra on the exactly same locations of τ for sample 1 and sample 2, we have the biggest mode, the second biggest mode, and the smallest mode sequentially. The first two spectra overlap in sample 1 and the last two spectra overlap a little in sample 2. The mode of the last spectrum from sample 1 is very small and this could be a

TABLE 4: Locations of τ for the modes of the spectra from the estimates of relaxation distribution f.

Experimental Data	Spectrum 1	Spectrum 2	Spectrum 3
Sample 1	$e^{-7.32}$	$e^{-5.90}$	$e^{-4.05}$
Sample 2	$e^{-8.42}$	$e^{-6.40}$	$e^{-4.97}$
Sample 3	$e^{-4.55}$	$e^{-2.03}$	$e^{-1.03}$
Sample 4	$e^{-3.55}$	$e^{-1.11}$	$e^{-0.01}$

spurious smoothing bump. So, the last spectrum in sample 1 might be artificial.

Using a GE Omega CSI system at a proton frequency of 85 MHz, water-saturated Texas Cream limestone samples were also analyzed. In this case, $\text{GCV}_C(\lambda)$ and $\hat{P}_C(\lambda)$ also provide us with very similar values of λ, which are very close to 0 as you can find in table 3. In figure 13, we present the estimates of the magnetization function and the relaxation distribution, where $\hat{\lambda}$'s are based on $\text{GCV}_C(\lambda)$. The values of $\hat{\lambda}$ are listed in table 3. With these samples, we also have three modes of spectra. The last mode of spectrum from sample 4 is very big which shows some different pattern of spectra from sample 1-3. See table 4 for the locations of τ for the modes of spectra. We do not find any spurious smoothing bumps from these samples.

Our laboratory magnetization mean function data sets are composed of measurements with very small errors. Considering the very low effect of noises, the estimate of relaxation distributions may be very accurate. However, the data sets in this dissertation had infinitesimal sizes of errors. Often, very small noises direct the values of smoothing parameters to near zero values. In this case, we doubt if $GCV(\lambda)$ or $\hat{P}(\lambda)$ is a good choice for selecting the smoothing parameter since this kind of problem may be viewed as a numerical approximation problem rather than a statistical problem.

We can use our estimates of the relaxation distribution to characterize the number of components for bulk porosity and fluid saturations and pore size distribution that characterize pore structures and fluid distribution in porous media, see Liaw et al. (1996) for detail.

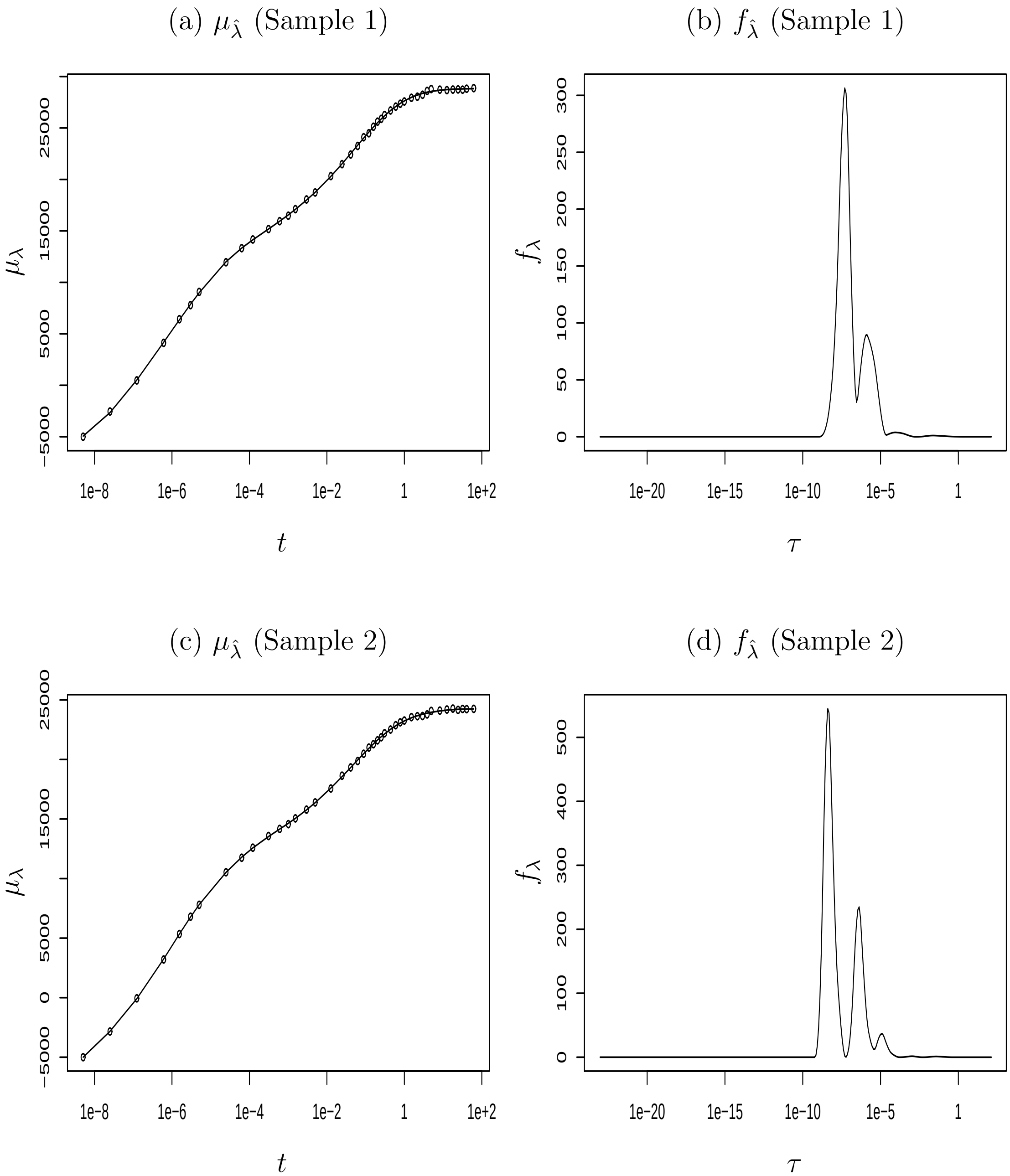

FIGURE 12: Estimates of magnetization function μ and relaxation distribution f for sample 1 and sample 2 from Opalinus clay (Benken, Switzerland). Dots are observed values of magnetization. Solid lines are estimates.

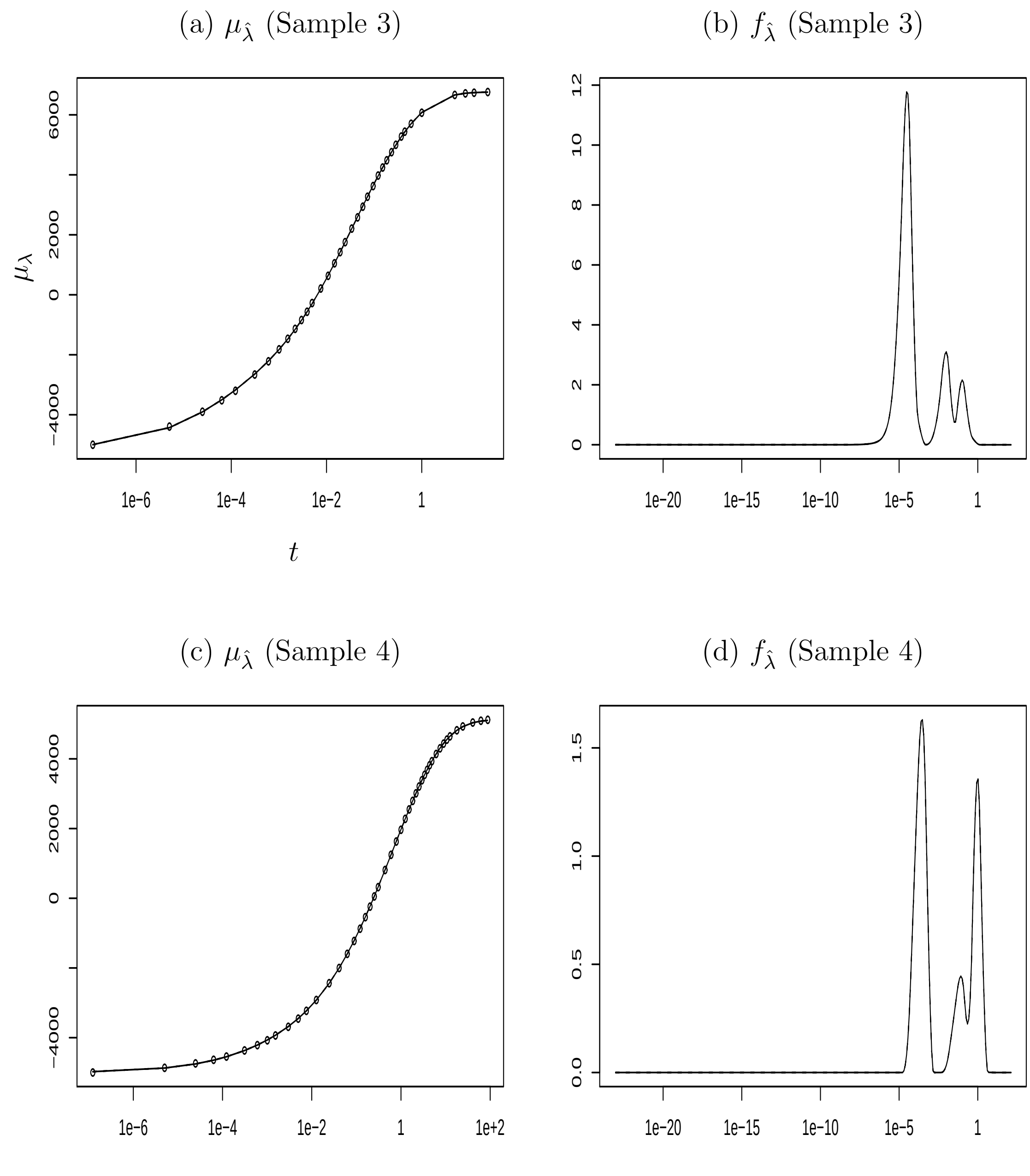

FIGURE 13: Estimates of magnetization function μ and relaxation distribution f for sample 3 and sample 4 from water-saturated Texas cream limestone. Dots are measured data. Solid lines are the estimates.

CHAPTER V

CONCLUSIONS

5.1 Summary of Results

The main objectives of this dissertation are the development of theoretical and methodological results for approximate MOR estimators. In Chapter II, we present an asymptotic development that relates the levels of smoothing for estimating a mean function μ and an associated inverse function f. Our developments are based on Fourier analysis. One of the main purposes of developing these theories is to find when the ill-posed inverse problem can be solved successfully using an optimal smoothing parameter that is related to the information about the mean function, i.e., one that can be estimated from the data at hand. We find that the conditions depend on the smoothness of the inverse function and the kernel function.

In Chapter III, stable and efficient computational algorithms for calculating the approximate MOR estimator are discussed in detail. The computational algorithms can be applied to an approximate MOR estimator with inequality, equality constraints or without constraints on the inverse function. We prove that the approximate MOR estimator is a solution for the unique global minimum of a minimization criterion for the problem. When inequality constraints are imposed on the inverse function, the parameter space is restricted. In this problem setting, data-driven automatic criteria $\mathrm{GCV}_C(\lambda)$ and $\hat{\mathrm{P}}_C(\lambda)$ for selecting the smoothing parameter are developed for the inequality constrained smoothing splines. We extend Bayesian credible intervals for the approximate MOR estimator and present a constrained version of the Bayesian intervals that are suitable for our problem setting.

A number of simulation studies are conducted for empirical investigations of the

$GCV_C(\lambda)$ criterion and of the Bayesian credible intervals. When the frequencies of the mean function reflect those of the inverse function, $GCV_C(\lambda)$ can provide very accurate estimates of the inverse function and the mean function, simultaneously. In those cases, we observe that the $100(1-\alpha)\%$ Bayesian credible intervals have very reasonable coverage for the inverse function. In most cases, whether the problem is highly ill-posed or not, the Bayesian intervals have approximately $100(1-\alpha)\%$ coverage for the mean function.

5.2 Future Research

There are several topics to be considered beyond the works completed in this dissertation. We now discuss some possible future research issues.

We have developed asymptotic theories focusing on a particular MOR estimator for the case where the linear functionals are defined by a convolution type of operator of the form

$$L_t f = \int_0^1 \frac{1}{2}\left[K(t-s) + K(t+s)\right] f(s)\ ds.$$

Similar developments are under consideration for a general convolution problem of the form

$$L_t f = \int_0^1 K(t-s) f(s)\ ds,$$

using the general estimation criterion

$$\sum_{i=1}^{n} (y_i - L_i f)^2 + \lambda \int_0^1 f^{(m)}(s)\ ds.$$

Extensions here might be possible using a parallel of the Demmler-Reinsch natural spline basis under proper assumptions on the corresponding eigenvalues.

While we have routinely applied pseudo-residuals suggested by Gasser et al. (1986) to estimate the σ^2 parameter in $\hat{P}(\lambda)$ and Bayesian credible intervals, variance estimation is also a very important problem to be considered in the MOR setting.

Some more investigations on estimation of σ^2 and the development of possible new estimation methods pose interesting problems for future research.

While most of our algorithmic work has focused on linear constraints in our model, developments with nonlinear constraints (or some other general constraints) parallel to the works in Chapter III might also be possible. Results of this nature would have implications for data-driven smoothing parameter selection and the Bayesian credible intervals in the presence of nonlinear constraints.

Other possible extensions include the development of boundary correction methods and applications to non-integral operators. Finally, in our model, we have assumed that the errors are uncorrelated random errors with common variance σ^2. It may be desirable to develop some methodology for correlated random errors.

REFERENCES

Aki, K. and Richards, G. (1980). *Quantitative Seismology: Theory and Methods.* San Francisco: Freeman.

Allen, D. (1974). The relationship between variable selection and data augmentation and a method of prediction. *Technometrics* **16**, 125–127.

Banavar, J. R. and Schwartz, L. M. (1989). Probing porous media with nuclear magnetic resonance in molecular dynamics in restricted geometrics. In *Molecular Dynamics in Restricted Geometrics*, 273–310. New York: Wiley.

Belton, P. S. and Ratcliffe, R. G. (1985). Nmr and compartmentation in biological tissues. *Progress in Nuclear Magnetic Resonance* **17**, 241–279.

Beyer, W. H. (1991). *CRC Standard Mathematical Tables and Formulae, 29th Edition.* Boston: CRC Press.

Bolt, B. A. (1980). What can inverse problems do for applied mathematics and the sciences? *Search* **11**, 194–198.

de Boor, C. (1978). *A Practical Guide to Splines.* New York: Springer-Verlag.

Brownstein, K. R. and Tarr, C. E. (1977). Spin-lattice relaxation in a system governed by diffusion. *Journal of Magnetic Resonance* **26**, 17–24.

Cox, D. D. (1988). Approximation of method of regularization estimators. *The Annals of Statistics* **16**, 694–712.

Cox, D. D. and O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *The Annals of Statistics* **18**, 1676–1695.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik* **31**, 377–403.

Davies, S. and Packer, K. J. (1990). Pore-size distribution from nuclear magnetic spin-lattice relaxation measurements of fluid-saturated porous solid. *Journal of Applied Physics* **67**, 3163–3176.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing.* New York: Marcel Dekker.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* **19**, 1257–1272.

Gallegos, D. P. and Smith, D. M. (1988). An NMR technique for the analysis of pore structure: Determination of continuous pore distribution. *Journal of Colloid and Interface Science* **122**, 143–153.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73**, 625–633.

Gill, P., Gould, N., Murray, W., Saunders, M., and Wright, M. (1982). Range space methods for convex quadratic programming. Tech. rep., SOL 82-14, Systems Optimization Laboratory, Stanford University, Dept. of Operations Research.

Golub, G., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–223.

Gore, J. C., Brown, M. S., Zhong, J., and Armitage, I. M. (1989). Prediction of proton relaxation rates from measurements of deuterium relaxation in aqueous systems. *Journal of Magenetic Resonance* **83**, 246–251.

Grunbaum, F. A. (1975). Remark on the phase problem in crystallography. *Proceedings National Academy of Science U.S.A.* **72**, 1699–1701.

Hart, J. and Yi, S. (1998). One-sided cross-validation. *Journal of the American Statistical Association* **93**, 620–631.

Inch, W. R., McCredie, J. A., Knispel, R. R., Thompson, R. T., and Pintar, M. M. (1974). Water-content and proton spin relaxation-time for maligant and nonmaligant tissues from mice and humans. *Journal of the National Cancer Institute* **52**, 353–356.

Kleinberg, R. L. (1996). *Encyclopedia of Nuclear Magnetic Resonance*, Vol. 8. Chichester: Wiley.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 481–492.

Lawson, C. L. and Hanson, R. J. (1974). *Solving least squares problems*. Englewood Cliffs, New Jersey: Prentice-Hall.

Liaw, H.-K., Kullkarni, R. N., Chen, S., and Watson, A. T. (1996). Characterization of fluid distribution in porous media by NMR techniques. *AIChE* **42**, 538–546.

Lukas, M. (1980). Regularization. In *Numerical Solution of Integral Equations*, pp. 151–182. Groningen: Sijthoff and Noordoff.

Mallows, C. L. (1973). Some comments on c_p. *Technometrics* **15**, 661–675.

Nychka, D. (1983). Solving integral equations with noisy data: An application of smoothing splines in pathology. *Proceedings in Computer Science and Statistics* **14**, 157–163.

Nychka, D. (1984). Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections. *Journal of the American Statistical Association* **79**, 832–846.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association* **83**, 1134–1143.

Nychka, D. (1990). The average posterior variance of a smoothing spline and a consistent estimate of the average squared error. *The Annals of Statistics* **18**, 415–428.

Nychka, D. and Cox, D. D. (1989). Convergence rates for regularized solutions of integral equations from discrete noisy data. *The Annals of Statistics* **17**, 556–572.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science* **1**, 502–527.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1986). *Numerical Recipes in Fortran* (Second edition). Cambridge: Cambridge University Press.

Rice, J. A. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics* **12**, 1215–1230.

Rice, J. A. (1986). Choice of smoothing parameter in deconvolution problems. *Contemporary Mathematics* **59**, 137–151.

Schumaker, L. L. (1990). *Spline Functions: Basic Theory.* New York: John Wiley & Sons.

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters* **9**, 229–235.

Tikhonov, A. (1963a). Regularization of incorrectly posed problems. *Soviet Mathematics Doklady* **4**, 1624–1627.

Tikhonov, A. (1963b). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady* **4**, 1035–1038.

Timur, A. (1969). Pulsed nuclear magnetic resonance studies of porosity, movable fluid, and permeability of sandstones. *Journal of Petroleum Technology* **21**, 775–786.

Villalobos, M. and Wahba, G. (1987). Inequality-constrained multivariate smoothing splines with application to the estimation of posterior probabilities. *Journal of the American Statistical Association* **82**, 239–248.

Wahba, G. (1980). Ill posed problems: Numerical and statistical methods for mildly, moderately and severely ill posed problems with noisy data. Tech. rep., Dept. of Statistics, University of Wisconsin.

Wahba, G. (1982). Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine. In *Statistical Decision Theory and Related Topics III*, Vol. 2, S.S. Gupta and J.O. Berger (eds), 383–418. New York: Academic Press.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society Ser. B* **45**, 133–150.

Wahba, G. (1986). Comments on a statistical perspective on ill-posed inverse problems. *Statistical Science* **1**, 521–522.

Wahba, G. (1990). *Spline Models for Observational Data.* CBMS-NSF series. Philadelphia: SIAM.

Watson, A. T. and Chang, C. T. P. (1997). Characterizing porous media with NMR methods. *Progress in Nuclear Magnetic Resonance Spectroscopy* **31**, 343–386.

Whittall, K. P. (1991). Investigation of analysis techniques for complicated NMR relaxation data. *Journal of Magnetic Resonance* **95**, 221–234.

VITA

Sang-Joon Lee earned a Bachelor of Science degree in statistics from Inha University in Inchon, Republic of Korea in 1991. He received a Master of Science degree in statistics under the direction of Dr. Woochul Kim from Seoul National University in Seoul, Republic of Korea in 1993. He received his Ph.D. under the direction of Dr. Randall L. Eubank in the Department of Statistics at Texas A&M University in May 2004. He is currently working in the Department of Biostatistics at M.D. Anderson Cancer Center, The University of Texas. He lives in Katy, Texas with his wife Sun-Ah Kim, daughters Alicia Kayoung and Sophia Jooyoung.

He can be contacted at:

Department of Statistics

Texas A&M University

3143 TAMU

College Station, TX 77843-3143